Segmentation and Classification of Heart Sound Signals

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Alex Paul Kamson

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, INDIA

MARCH 2022

Certificate

This is to certify that the thesis entitled "**Segmentation and Classification of Heart Sound Signals**", submitted by **Alex Paul Kamson** (146102024), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under our supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. S. Dandapat,
Professor
Dept. of Electronics and Electrical Engg.
Indian Institute of Technology Guwahati
Guwahati - 781039, India.

Dated:
Guwahati.

Dr. L. N. Sharma,
Senior Technical Officer
Dept. of Electronics and Electrical Engg.
Indian Institute of Technology Guwahati
Guwahati - 781039, India.

Dated:
Guwahati.

To

My beloved parents

Joseph Kamson and Regina Kamson

for their blessings, love and encouragements

My guide

Prof. S. Dandapat and Dr. L. N. Sharma

for their guidance, support and inspiration

&

My sisters

Jenny Kamson and Ermina Kamson

for their love and support

Acknowledgements

First of all, I would like to express my gratitude to God for enabling me to complete my thesis work.

I would like to express my indebtedness appreciation to my research supervisors, Prof. S. Dandapat and Dr. L. N. Sharma, for providing me with an opportunity to work under their guidance. Their constant guidance and inspiring advice have helped me in every stage of my research work. Their suggestions were crucial in making my work as flawless as possible. It would not have been possible to complete this thesis without their sincere efforts and motivations.

I am thankful to my doctoral committee members, Prof. P. K. Bora, Prof. R. Sinha, Prof. S. R. M. Prasanna, and Dr. Salil Kashyap, for their encouragement and valuable suggestions on my work. I am very grateful to them for their insightful comments and constructive criticisms on the work to bring it to the current form.

I sincerely thank Prof. P. K. Bora and Prof. S. Dandapat for their valuable advice and motivations at difficult stages of my academic life. Their words were my inspiration to push me through hardships.

I would like to thank my seniors, Dr. Banriskhem K. Khonglah, Dr. Jiss J. Nallikuzhy, Dr. Suman Deb, and Dr. Rajesh K. Tripathy, for their help and support.

I am thankful to my friends Dr. Sisir, Sarfaraz, Dr. Vivek, Vijit, Dr. Kaushik, Ato, Moakala, Angelus, E. Prabhakararao, Vineeta Das, Samarjeet, Sibasis, and Debasish for their valuable suggestions and help in shaping my work. My sincere gratitude to all the research scholars in Signal and informatics Lab and Electro-Medical and Speech Technology (EMST) for their help and support during my research work.

Finally, my heartiest thanks to my parents and family members for their constant blessings and silent prayers for the successful completion of my thesis work.

Abstract

Phonocardiogram (PCG) is a graphical representation of a heart sound signal. It provides information about the myocardium, septum and valvular events, resulting as a direct consequence of haemodynamic and heart rhythm. Any dysfunction of the heart is reflected as an abnormal sound in the PCG signal. To date, the auscultation of the heart is used as an initial screening step for further diagnosis. But with a growing repository of information on pathological conditions, there is a need for a computer-aided diagnostic (CAD) system for accurate detection of cardiac abnormalities. This thesis documents our investigation on noisy PCG signals for segmentation and classification of heart sound signals.

The first part of the thesis introduces a hybrid dual filtering method to denoise PCG signals. The hybrid method combines either band-pass filter (BPF) or wavelet transform (WT) with the total variation filter (TVF). It works by first passing a noisy PCG signal through the BPF/WT filter to remove high-frequency noise components that do not overlap with the frequency band of the normal heart sound. Next, the output signal is passed through the TVF to suppress the low amplitude residual noise in the systole and diastole intervals. The quality of the resulting signal from TVF depends on the choice of the regularization parameter controlling the degree of the smoothing process. In the proposed TVF, the regularization parameter is estimated by a data-dependent method that computes the complexity of the signal as the sample entropy. Knowing the complexity of signals helps in conditioning the filtering process from over smoothing the signal waveform. The overall performance of the proposed method is better than the existing methods.

In the second part, the logistic function is investigated to enhance the envelope peaks due to the S1 and S2 sounds. Traditionally, Shannon entropy and Shannon energy are employed as a medium to emphasize the PCG envelope. But they are sensitive to noise and may result in noisy envelope features. The logistic function amplitude moderation (LFAM) method is presented as an improvement over the traditional methods. The LFAM maps the predominant amplitudes of FHS to the upper asymptote of the S-curve and the noise signals to its asymptotic tail by adjusting the growth rate and the location of the curve mid-point. The main contribution is calibrating these parameters according to the critical upper amplitude (x_{uc}) and the lower amplitude (x_{lc}) corresponding to the level above which the signals are categorized as loud sound and below which the signals will be discarded as weak noise signals, respectively. Then their optimal values are regressively obtained for each signal by histogram analysis of signal intensity distribution. The x_{uc} is calculated as the amplitude that maximizes the number of signal samples in the category of loud sound. And the x_{lc} is calculated as the amplitude that maximizes the number of signal samples in the weaker sound level while preserving maximum signal information in the upper amplitudes. The transformed waveform enhances the FHS envelop peaks uniformly and provides better segregation from the systole and diastole intervals by a large amplitude margin. Compared to the traditional approaches, the proposed method significantly improves the envelope feature to identify the FHS.

The third part of the thesis presents a multi-mode diastolic duration model for the hidden semi-Markov model (HSMM) based heart sound segmentation (HSS). In the proposed duration model, each mode represents the clusters of duration values obtained under permissible variations by the hierarchical agglomerative clustering method. The intuition is that multiple peak distribution yields a sharper likelihood gradient around the expected mode resulting in more accurate discrimination. Each mode acts as a local reference for the HSMM to determine the hidden state

and adjust the corresponding duration based on the maximum likelihood criterion. The modification improves the versatility of the classifier to handle the PCG signal with variable heart rate and improves the overall performance.

The fourth part of the work investigates the potential of the proposed subband energy (SBE) feature and the inter-segment spectral correlations to classify the PCG signal into a normal, murmur, and noisy categories. It also investigates the potential of existing features, including log-magnitude spectrogram, SBE defined over bark scale, Mel-frequency cepstral coefficients (MFCCs), and the rhythm pattern (RP). The proposed SBE feature is specified based on the primary frequency components of each class of the heart sound. And, the inter-segment correlations are measured to investigate the changes in different heart sound intervals. On using individual features for the classification of PCG signal, the proposed SBE feature performs better than the others. However, when tested with different combinations of features, the inter-segment correlation between systole and diastole pair and 1-4 order MFCC feature performs better than the rest. From the study, it is convinced that the proposed features have the potential for improvement. With a more elaborate analysis, the features may be extended to classify murmurs and pathology.

Keywords: Phonocardiogram (PCG), fundamental heart sound (FHS), total variation filter (TVF), logistic function amplitude moderation (LFAM), Shannon entropy or Shannon energy mode selection (SE2MS), hidden semi-Markov model (HSMM), heart sound segmentation (HSS).

Contents

List of Figures

List of Tables

List of Abbreviations

AC	Air Conditioner
AD	Aortic Diseases
AR	Aortic Regurgitation
AS	Aortic Stenosis
AV	Atrioventricular
AWGN	Additive White Gaussian Noise
BiGRNN	Bidirectional Gated Recurrent Neural Network
BiLSTM	Bidirectional Long-Short Term Memory
BPF	Band Pass Filter
bpm	Beats per minute
CADS	Computer-Aided-Diagnostic System
CVD	Cardiovascular Diseases
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FHS	Fundamental Heart Sound
FN	False Negative
FP	False Positive
GRNN	Gated Recurrent Neural Network
HCD	Heart Cycle Duration

hPDF	Histogram based Probability Density Function
HMM	Hidden Markov Model
HRV	Heart Rate Variation
HSC	Heart Sound Classification
HSMM	Hidden Semi-Markov Model
HSS	Heart Sound Segmentation
IMFs	Intrinsic Mode Functions
LFAM	Logistic Function Amplitude Moderation
LogMS	Log Magnitude Spectrogram
LPF	Low Pass Filter
LR	Logistic Regression
LSTM	Long-Short Term Memory
LTI	Linear time Invariant
MAD	Median Absolute Deviation
MAP	Maximum a Posteriori
MFCC	Mel-Frequency Cepstral Coefficients
MITHSDB	Massachusetts Institute of Technology Heart Sound Database
MM	Majorization-Minimization
MPC	Miscellaneous Pathological Conditions
MR	Mitral Regurgitation
MS	Mitral Stenosis
MVP	Mitral Valve Prolapsed
OGS	Overlapping Group Sparsity
OGS-TVF	Overlapping Group Sparsity-Total Variation Filter
PCG	Phonocardiogram
PDF	Probability Density Function
PR	Pulmonary Regurgitation
PS	Pulmonary Stenosis

PSD	Power Spectral Density
RTAN	Real Time Ambient Noise
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SBE	Sub-Band Energy
SE2MS	Shannon Entropy or Shannon Energy Mode Selection
SFER	Signal-to-Filter Error ratio
SNR	Signal-to-Noise Ration
STFT	Short-Time Fourier Transform
SVD	Singular Value Decomposition
TKE	Teager Kaiser Energy
TN	True Negative
TP	True Positive
TV	Total Variation
TVF	Total Variation Filter
VHD	Valvular Heart Diseases
VSD	Ventricular Septal Defect
WPT	Wavelet Packet Transform
WT	Wavelet Transform

List of Symbols

A2	Heart sound components due to the closure of the aortic valve
$a_{i,j}$	Probability of transition from the state i to state j
a_m, $a_m(k)$	Approximation coefficients of m^{th} level decomposition
α	Growth rate of the sigmoid curve
$\grave{\alpha}$	Shape parameter of Gamma prior
β, β_1, β_2	Shifting parameter of the logistic function
$\grave{\beta}$	Scale parameter of Gamma prior
$b_j(O_t)$	Probability of observation at time t in the state j
c_i	Centroid of the cluster i
d	Duration
$\mathbf{D}$	First order derivative matrix
$\text{DF}_{\text{high\_amp}}(i)$	Distribution function of signals above intensity level i
$\text{DF}_{\text{low\_amp}}(i)$	Distribution function of signals below intensity level i
$\text{DF}_{\text{SoA}}(i)$	Distribution function of the relative sum of the amplitude above intensity level i
$\text{DF}_{\text{loudness}}(i)$	Distribution function of the loudness level for signal above intensity level i
Dia	Diastole segment
d_m, $d_m(k)$	Detail coefficients of m^{th} level decomposition
d_{S1}	Duration of S1
d_{S2}	Duration of S2
d_{siDia}	Duration of diastole without S2
d_{siSys}	Duration of systole without S1
$\delta_t(i)$	Probability of the observation sequence $\{O_1, ..., O_t\}$ being in state i at time t

$\Delta_t(j)$	Parameter tracking the likely duration of being in state j when calculated at the time instance t
ed_{sys}	Estimated duration of the systole
E_{dB}	Phon/Decibel scale
E_{Sone}	Bark scale sonogram
$e(t)$	Slow varying component or envelope of a signal
f	Frequency
$f(*)$	Filter output
$\hat{f}_n$	Normalization factor of denoised signal
$F(\mathbf{y})$	Objective function of TVF
g	Analysis low pass filter coefficients
$\tilde{g}$	Synthesis low pass filter coefficients
$G(\mathbf{y})$	Convex majorizer of the objective function in TVF
γ	Hyperparameter involved in the OGS penalty function
h	Analysis high pass filter coefficients
$\tilde{h}$	Synthesis high pass filter coefficients
Hilbert_E	Hilbert envelope
HEoDF	Homomorphic envelope after dual filtering
Homo_E	Homomorphic envelope
$\text{hPDF}_{5\%}$	Histogram probability density function of signal whose intensity level is below 5% of maximum value
$\text{hPDF}_{5-35\%}$	Histogram probability density function of signal whose intensity level is between 5% and 35% of maximum value
$\text{hPDF}_{35\%}$	Histogram probability density function of signal whose intensity level is above 35% of maximum value
$\mathbf{I}$	Identity matrix
iEnergy	Energy of the signal at intermediate intensity levels
k	Integers

K	Group size
L	Length of signal sequence
λ	Regularization tuning parameter or smoothing parameter of TVF
Λ_i	Diagonal matrix at i^th iteration
LFAM_E	Logistic function amplitude moderation based envelope
$\hat{\Lambda}$	HMM or HSMM model
Λ_i	Diagonal matrix at i^{th} iteration
$L(\mathbf{y})$	Objective function of TVF using MAP estimation
M	Total number of wavelet decomposition levels
M1	Heart sound components due to the closure of the mitral valve
M_Acc	Average of sensitivity and specificity
M_signal	Moderated signal
ms	Millisecond
$\mu(*)$	Mean or average value
n	Integers
N	Length of signal sequence
$\hat{n}_c$	Number of clusters
$\hat{n}_i$	Number of data points in the cluster i
O	Observation sequence
$o(t)$	Fast oscillating components of a signal
O_t	Observation at time t
P2	Heart sound components due to the closure of the pulmonary valve
ϕ_m	Scaling Function of WT
Φ^K	Probability of template pairs of size K that is matching
π	Initial state probability distribution of HMM/HSMM
$p_i(d)$	Duration probability of being in state i
p_SBE$_{log}$	Proposed sub-band energy from log-magnitude spectrogram
p_SBE$_{lφg}$	Proposed sub-band energy from the normal magnitude spectrogram

ψ_m	Wavelet Function of WT
$\Psi_t(j)$	Parameter tracking the most likely past state that maximizes the transition to the present state j at time t
P_+	Positive predictivity or precision
Q	State sequence
$\hat{Q}$	Best state sequence
r	Noise filter parameter
s	Second
S	Set of all possible states
S1	First heart sound
S2	Second heart sound
S3	Third heart sound
S4	Fourth heart sound
SampEn	Sample entropy
Se	Sensitivity
SE2MS_E	Shannon entropy or Shannon energy mode selection based envelope
SEng_E	Shannon Energy based Envelope
SEnt_E	Shannon Entropy based Envelope
siDia	Diastole segment excluding the S2 sound segment
siSys	Systole segment excluding the S1 sound segment
$\sigma(x)$	Logistic function
Sp	Specificity
Sys	Systole segment
t	Time
T1	Heart sound components due to the closure of the tricuspid valve
θ	A constant
$\varphi(\mathbf{x})$	Penalty function
ϱ^2	Variance

$\varrho_{c_i}^2$	Variance of the cluster i
w	weights
x_0	Mid-point of the sigmoid curve
$\mathbf{x}$, $x(n)$	Input signal sequence
x_{lc}	Lower cut-off amplitude
x_{uc}	Upper cut-off amplitude
$\mathbf{x}_{n,K}$	K-point vector of $\mathbf{x}$ starting from any instance n
$\mathbf{y}$, $y(n)$	Output signal sequence
$\bar{\mathbf{y}}_i$	Convergent sequence at i^{th} iteration

1

Introduction

Contents

This thesis documents our investigation on the analysis of heart sound signals to assess the heart and related pathology. Heart sound is a vibrational sound originating from the mechanical activity of the heart. Being one of the most studied cardiac signals, it provides valuable information for the detection of cardiovascular diseases (CVDs) [3, 8, 9]. A normal heart sound from a healthy subject has a pair of S1 and S2 sounds in every cardiac cycle. The normal heart rate ranges between 60 and 100 beats per min [10, 11] at resting condition. Anomalies such as the appearance of any additional sound other than S1 and S2 sounds, abnormally loud or soft sound level, or unusual variation in sound durations may indicate some cardiac ailments. Such anomalies originate from conditions such as conduction failure, dysfunction of heart valves, septal defects, or cardiomyopathy [12].

Studying the mechanism behind the production of heart sound has revealed that heart sound carries information that reflects the structural and functional integrity of the heart. Many heart diseases cause murmurs and aberrations in heart sound before they are realized in other signals [13]. Because of this reason, heart auscultation is used as a primary diagnostic tool for ages. But, effective diagnosis of heart diseases is limited by the ability of human auditory perception. Therefore, there is a need for a computer-aided-diagnostic (CAD) system that can analyze and interpret or visualize the complex heart sound signals and help diagnose the abnormalities of the heart. This Chapter introduces the heart sound signal and various cardiac ailments, including pathological characteristics of heart sound that may be useful for developing an automated CAD system.

1.1 The heart and function of the heart valves

The heart is a muscular organ that pumps blood throughout the cardiovascular system. A schematic cross-sectional view of the heart is shown in Fig. 1.1. The heart acts as two separate pumps: the right heart pumps blood to the lungs, and the left heart circulates blood to other parts of the body. Each of these heart pumps has two pulsatile chambers known as the atrium and ventricle. Each atrium acts as a weak primer pump that collects blood and

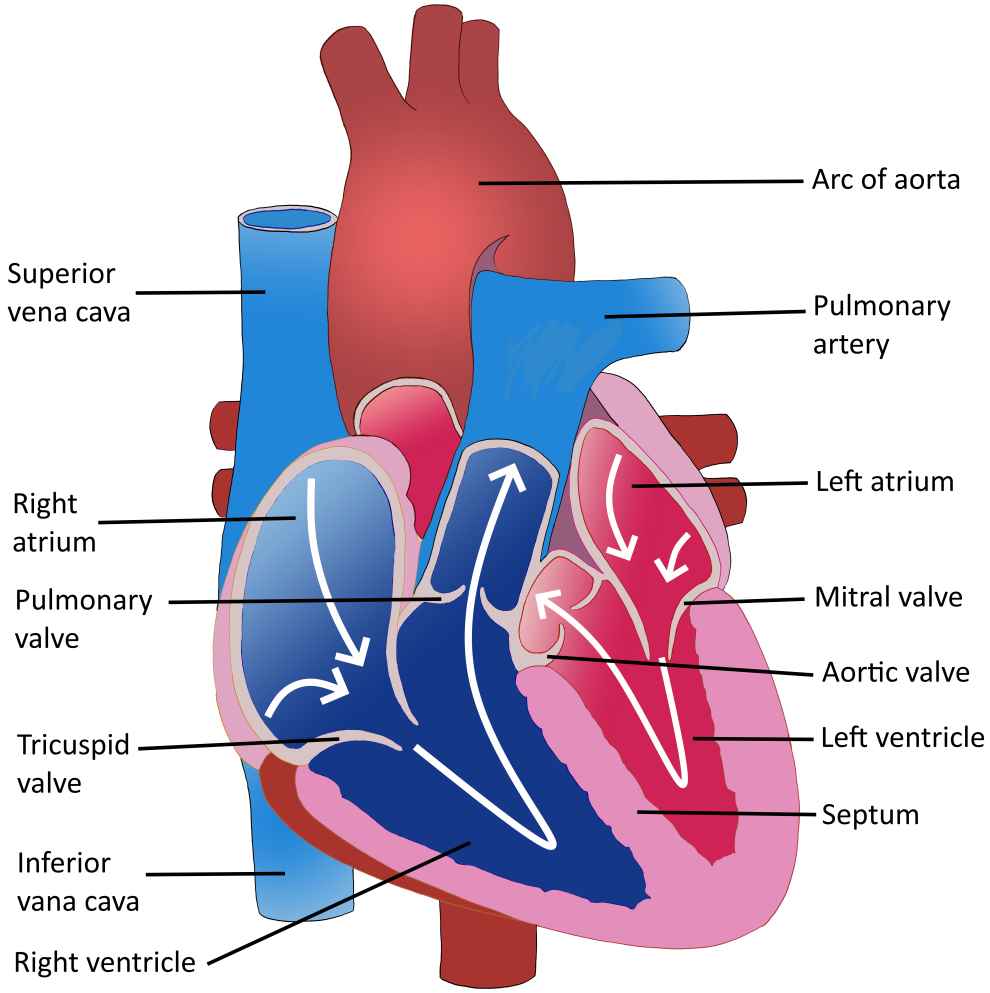

Figure 1.1: Internal structures of the heart showing all four chambers.

moves it into the ventricle. The ventricle is the main pumping force that propels the blood to either pulmonary circulation or systemic circulation. The mechanism of the heart is controlled by a continuous succession of electrical depolarization and repolarization of heart muscles called cardiac rhythmicity. During depolarization, the heart muscles contract and pumps the blood resulting in systole. During repolarization, also known as the diastole, the heart muscles undergo relaxation and refill the heart chambers with blood. There are four valves through which blood passes before leaving each heart chamber, namely: tricuspid valve, pulmonary valve, mitral valve, and aortic valve (see Fig. 1.1). They prevent the backflow of blood during systole, thereby maintaining the unidirectional flow inside the heart.

During ventricular systole, the atrioventricular (i.e. tricuspid and mitral) valves close, leading to the accumulation of blood in the right and left atria from the vena cava and pulmonary vein, respectively. As soon as the ventricular diastole starts, the semilunar valves (i.e. the aortic and pulmonary valves) close and prevent the backflow of blood from the aorta and pulmonary arteries. During this repolarization period, the atrioventricular valves remain open allowing blood from the atria to fill up the ventricles. In the next ventricular systole, blood from the right ventricle is pumped through the pulmonary circulation for oxygenation. Simultaneously, the left ventricle pumps the oxygenated blood through the arteries to different parts of the body. These cardiac events starting from the beginning of one heartbeat to the next heartbeat are called the cardiac cycle.

1.2 Normal heart sound and pathological changes in the sound

Heart sound consists of two audible sounds: ‘lub’ and ‘dub’ which are repeated after every cardiac cycle. These sounds are produced due to the snap shut of heart valves causing the vanes of the valves and the surrounding wall to vibrate. The opening of heart valves is a relatively slow process and it does not produce any sound.

1.2.1 First heart sound

The first heart sound (S1) or ’lub’ is produced when the atrioventricular valves close (tricuspid and mitral valves). The sound is generated as a consequence of the slapping together of the valve leaflets along with the vibrations of valves and enclosing walls of the heart chambers. An illustration indicating the instance of ventricular systole and the associate state of the heart valves is shown in Fig. 1.2. During ventricular systole, the heart muscle contraction builds up high pressure inside the chamber that causes sudden reflux of blood against the atrioventricular valves resulting in taut valves and chordae tendineae to bounce back and forth. This mechanism causes the valves and ventricular wall vibration and experiences the turbulence of blood, which propagates through the heart myocardium, and the chest wall,

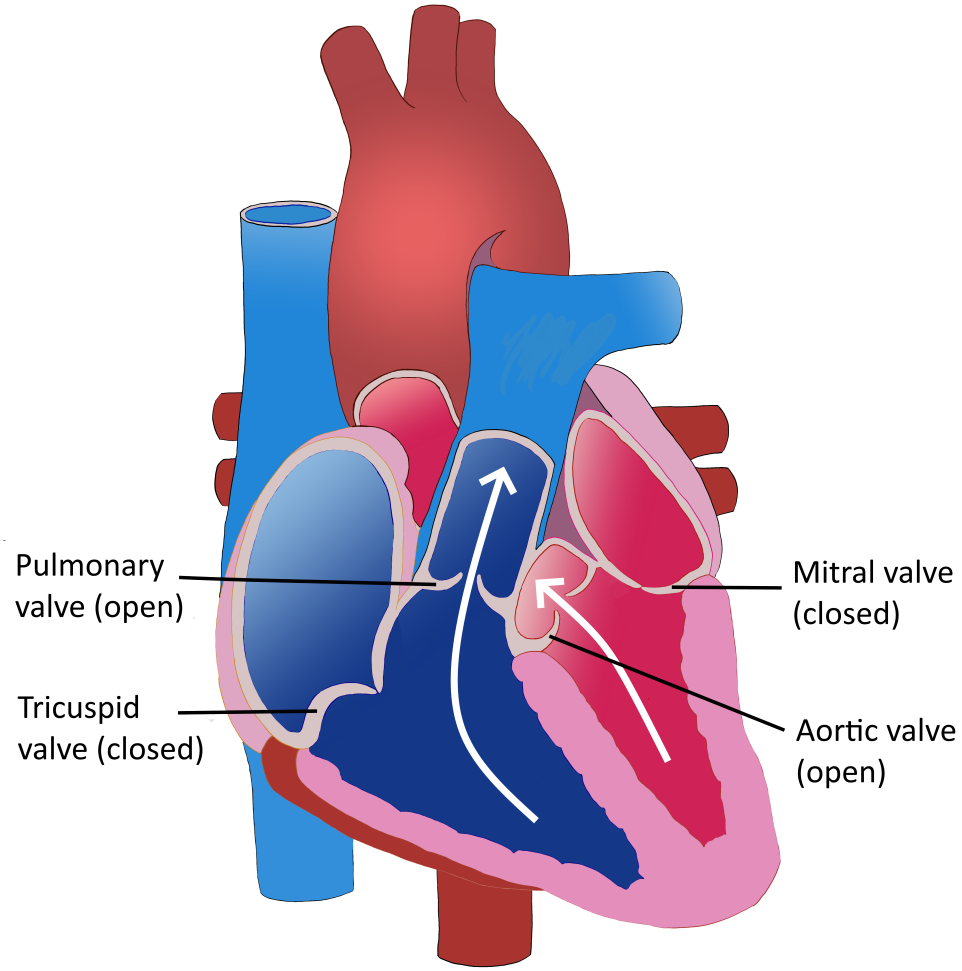

Figure 1.2: During ventricular contraction, the mitral and tricuspid valves close, preventing the backflow of blood to the atria.

where it is perceived as the first heart sound.

The auscultation of the heart is performed by placing the stethoscope diaphragm on the chest. Then move the stethoscope from one site to another, identifying the valve location from where the sound originates. The systematic approach of auscultation is by first placing the diaphragm over the aortic area (second right intercostal space). Then inched towards the pulmonary area (second left intercostal space), the tricuspid area (fourth left intercostal space at the sternal border), and the mitral area (cardiac apex, which is between the 5th and 6th intercostal spaces in the mid-clavicular line) [14, 15], as shown in Fig. 1.3. The sounds produced from all the valves are usually audible in all these sites. The experts distinguish

them by the process of elimination, which comes from experience.

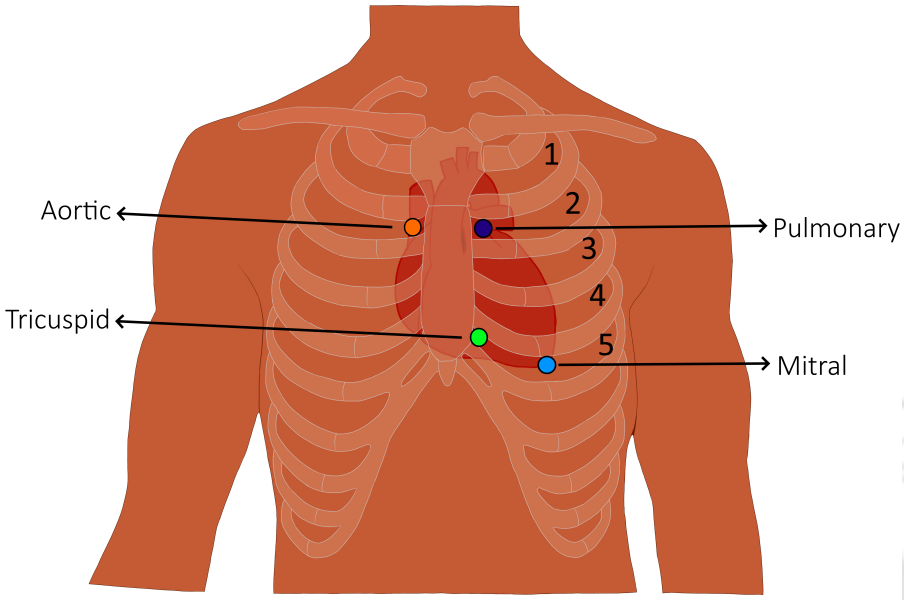

Figure 1.3: The four standard sites of heart sound auscultation

The first heart sound, S1, is associated with the closure of the mitral and tricuspid valves, and they are linked to the two audible components: M1 and T1 sounds, respectively. The M1 sound is much louder, has a higher frequency, and occurs earlier than the T1 sound. They are best heard at the apex of the heart. The two components of S1 sound are produced simultaneously as one whole sound. Sometime, a normal asynchrony may cause a delay of 20-30 msec between the closure of mitral and tricuspid valves (best heard during expiration). A split S1 sound may be heard along the tricuspid area where T1 sound might also be prominently audible, see Fig. 1.4.

Clinical significance

For clinical significance, parameters such as the quality, intensity, and degree of splitting of S1 sound are examined. These specifications respond differently to different anomalies. The intensity of S1 sound is associated with the rate of closing of atrioventricular valves, the pressure exerted during ventricular contraction, and the anatomical integrity of valves. For example, the S1 sound appears louder in mitral stenosis due to greater exertion of calcified

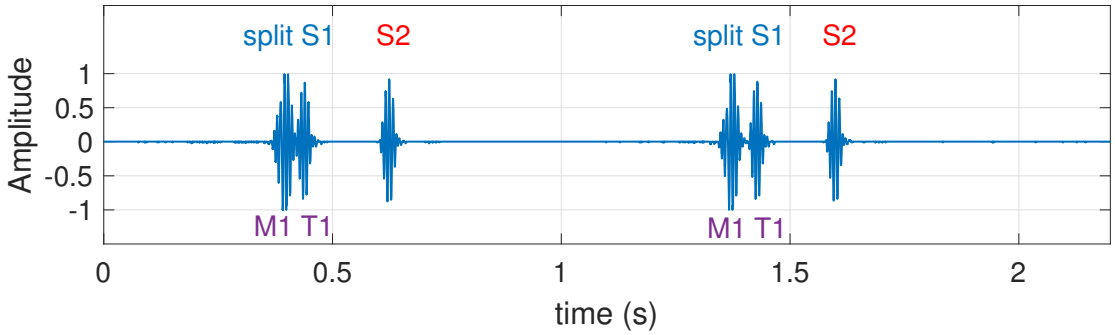

Figure 1.4: Split S1 sound indicating the M1 and T1 sound components.

mitral leaflets during closing. Certain heart conditions that increase myocardial contractilities, such as exercise, anxiety, anemia, fever, pregnancy, and thyrotoxicosis, may assert a louder S1 sound.

Soft S1 sound is common in the subject suffering from first-degree heart block. Other abnormalities that may reduce the intensity of S1 sound are mitral regurgitation, ventricular septal defect, and acute aortic regurgitation.

Wide splitting of S1 sound is always abnormal. The split may increase to 60 ms. It is often observed in patients with right bundle branch block, Ebstein's anomaly, or conditions with activation delay such as ventricular ectopic beats, ventricular tachycardia, atrioventricular block, and left ventricular pacing [15].

1.2.2 Second heart sound

The second heart sound (S2) or 'dub' is produced when the semilunar valves close. At the end of the systole, the ventricles start repolarizing, and the semilunar valves block the backflow of blood from the arteries. The elastic stretch of the semilunar valves recoils the blood back into the arteries producing short reverberation of blood between the heart walls and the valve leaflets. These vibrations propagate to the chest wall as the S2 sound. The instance of ventricular diastole and the associated state of the heart valves is shown in Fig. 1.5.

The S2 sound also has two audible components: A2 and P2 sounds associated with the

closure of aortic and pulmonary valves, respectively. After A2 sound, there may be a split delay of 20 to 80 ms before producing P2 sound [15]. An example of the split sound is shown in Fig. 1.6. In a healthy subject, a mild splitting of S2 sound occurs during inspiration and disappears on expiration. Splitting is best perceived when the stethoscope is placed within the second and third left intercostal space where P2 sound is confined.

Clinical significance

The clinical evaluation of S2 sound requires determining the split duration and the intensity relation of A2 and P2 sounds. Because the abnormalities in semilunar valves and the ventricles are reflected in S2 sound and subsequent diastolic intervals. Usually, the A2 sound is much louder and occurs before the P2 sound. During inspiration, P2 sound may be delayed by 20 to 80 ms. With expiration, there is hardly any delay between A2 and P2. Rarely it splits at most by 40 ms. It is abnormal if the S2 sound split's width exceeds 40 ms during expiration. Such occurrence is observed in a subject suffering from delayed activation of the right ventricle, atrial septal defects, pulmonary hypertension, pulmonary stenosis, and more.

Fixed splitting of S2 sound is another abnormal condition in which the split interval of A2 and P2 remains unaffected by respiration. This symptom may associate with atrial septal defects showing pulmonary vascular resistance or severe right heart failure.

A delay in the aortic valve closure may cause a condition known as the paradoxical or reverse split. Here, P2 precedes A2. It results from prolonged left ventricle activation. The symptom may be associated with left bundle branch block, hypertrophic cardiomyopathy, aortic stenosis, acute myocardial infarction, or left ventricular dysfunction.

1.2.3 Duration and pitch of the normal heart sound

The S1 and S2 sounds are short-duration sounds. These two sounds will be addressed as the fundamental heart sound (FHS) in the later section. The average duration of S1 sound is 0.12 seconds, with some tolerance of about 22 ms (0.022 s). The semilunar valves are tauter than the atrioventricular valves. Therefore, the S2 sound is slightly shorter than the S1 sound.

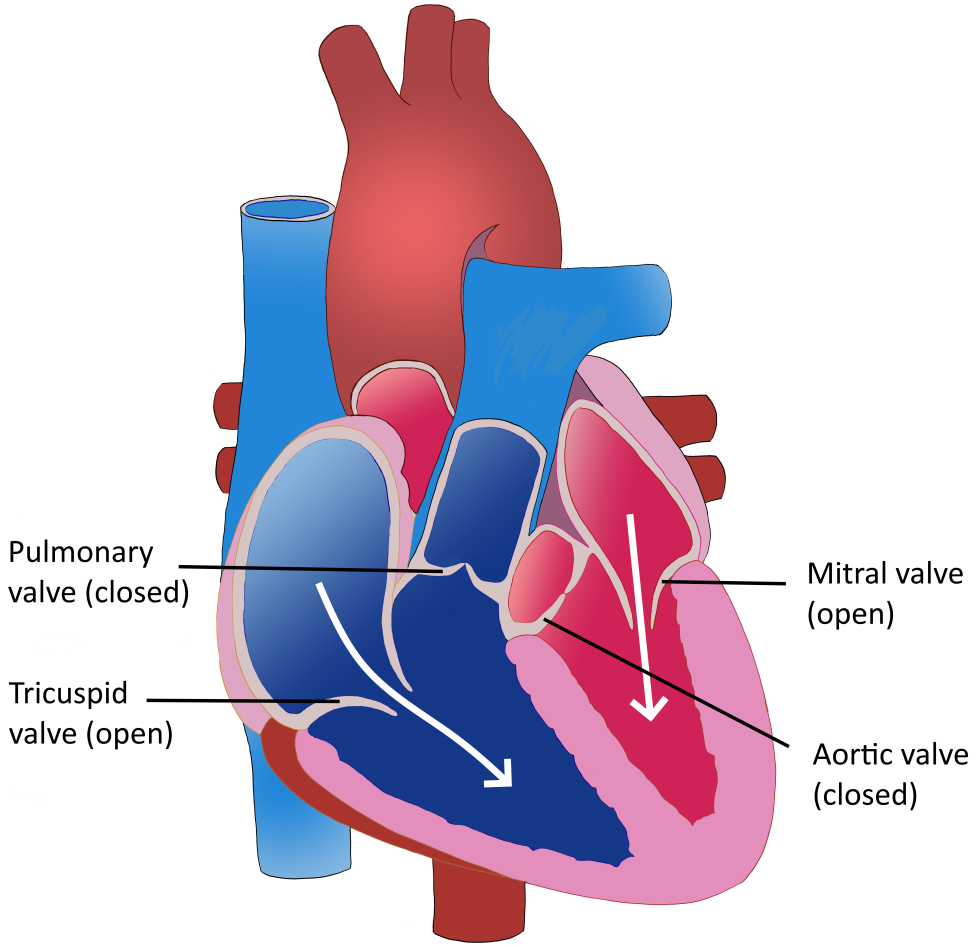

Figure 1.5: During ventricular relaxation, the Aortic and Pulmonary valves close, preventing the backflow of blood from the aorta and pulmonary artery to the ventricles.

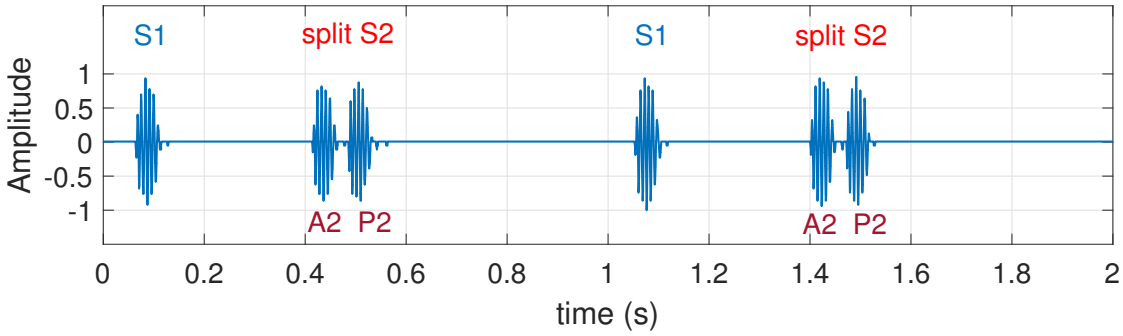

Figure 1.6: Split S2 sound indicating the A2 and P2 sound components.

It is about 92 ± 22 ms duration [16].

Every cardiac cycle has a pair of S1 and S2 sounds. The S1 sound marks the beginning of ventricular systole, and the S2 sound indicates the beginning of ventricular diastole.

Morphologically, they are identified based on the duration difference between the systole and diastole intervals. The duration of systole is always shorter than the duration of diastole. The total time duration of the cardiac cycle or heart cycle duration (HCD) is the reciprocal of the rate at which the heart beats. The heart rate (or the number of heart cycles per minute) for an average person is between 60-100 beats per minute (bpm) [17]. Analyzing the spectral properties of the heart sound signals have shown that the frequency range of S1 is 20-150 Hz (most signal energy concentrated in lower frequencies of 25-45 Hz) and S2 is 20-200 Hz (energy concentrated in 55-75 Hz) [5, 18–20].

1.2.4 Third and fourth heart sounds

Occasionally weak gallop sounds called S3 and S4 may appear in the diastole interval. The S3 sound is a very low amplitude and low-pitch sound that occurs in the early diastolic period after S2 sound ($\approx$ 120-150 msec after S2). It has a small frequency bandwidth of 25-50 Hz. Finding S3 sound is normal in subjects with high cardiac output, such as highly trained athletes and pregnant subjects. It can be pathologically associated with cardiac conditions such as mitral regurgitation, hypertrophic cardiomyopathy, and anemia. S4 is another low-pitch sound caused by rapid atrial contraction. It occurs at the late diastolic period just before the S1 sound ($\approx$ 90 msec before S1). It has a frequency composition ranging between 15 Hz and 65 Hz. The S4 sound is always pathological in origin. Due to close proximity with the FHS, the gallop sounds may be mistaken for split heart sounds. Even for a well-trained physicist, it is difficult to identify them.

1.2.5 Murmur

Auscultation of the heart tells much about the health and functional activities of the heart valves, surrounding myocardium, and septum. Murmur is one such clinically significant sound. The heart murmurs are rumbling sounds produced by the turbulent blood flow under high pressure when it forces through a narrow opening. Heart murmurs can be broadly categorized into stenosis and regurgitation. A stenosis murmur is a sound resulting from

the insufficient opening of heart valves. Due to the narrow opening of valves, the blood flow is restricted, causing vibrating turbulence of blood. A regurgitation murmur is produced when the heart valves do not close adequately. A leaky valve causes the backflow of blood through the narrow gap, which creates a rumbling sound. Not every murmur is associated with valve abnormalities [21]. Septal defects in children also produce murmur sounds. A temporary increase in blood flow may also cause murmurs, which is not pathological. The non-pathological murmurs are known as innocent murmurs.

There are two commonly found valvular heart diseases (VHD), namely Aortic stenosis (AS) and mitral regurgitation (MR). Examples of their phonocardiogram are shown in Fig. 1.7 (a) and Fig. 1.7 (b). A murmur due to ventricular septal defect (VSD) is shown in Fig. 1.7 (c). The nomenclature of the VHD is based on the name of the valve from where the murmur originates. Aortic stenosis (AS) is because of the insufficient opening of the aortic valve, and mitral regurgitation (MR) is because of the improper closing of the mitral valve. Similarly, there are other VHD such as pulmonary stenosis (PS), pulmonary regurgitation (PR), mitral valve prolapsed (MVP), etc.

Some clinically significant features that help categorize the murmurs are intensity (loudness), intensity configuration of murmurs, position and duration of murmurs in the cardiac cycle, frequency (pitch), and tonal effects of murmurs. The intensity of murmurs may range from grade 1 (or faintly audible) to grade 6 (or extremely loud murmur audible without contact between the stethoscope and the chest wall). The intensities are graded using the Levine scale [22]. The intensity configuration of a murmur plotted with respect to time may form a shape that resembles a crescendo, decrescendo, or crescendo-decrescendo. The intensity configuration depicts the blood pressure building up inside the heart chambers and the severity of the dysfunctions. For example, the murmur of mitral regurgitation is typically decrescendo with variable intensity in systole. However, mitral regurgitation associated with papillary muscle dysfunction is crescendo–decrescendo intensity in mid systole [23]. The murmur of aortic stenosis typically has a crescendo–decrescendo intensity with low to medium frequency ranges.

Murmurs manifest in diverse frequency ranges depending on the severity and nature of the pathology. The frequency composition may go as high as 700 Hz [24]. The frequency specs of different heart sound signals are tabulated in Table 1.1 below. Some typical symptoms associated with commonly occurring CVD are also tabulated in Table 1.2.

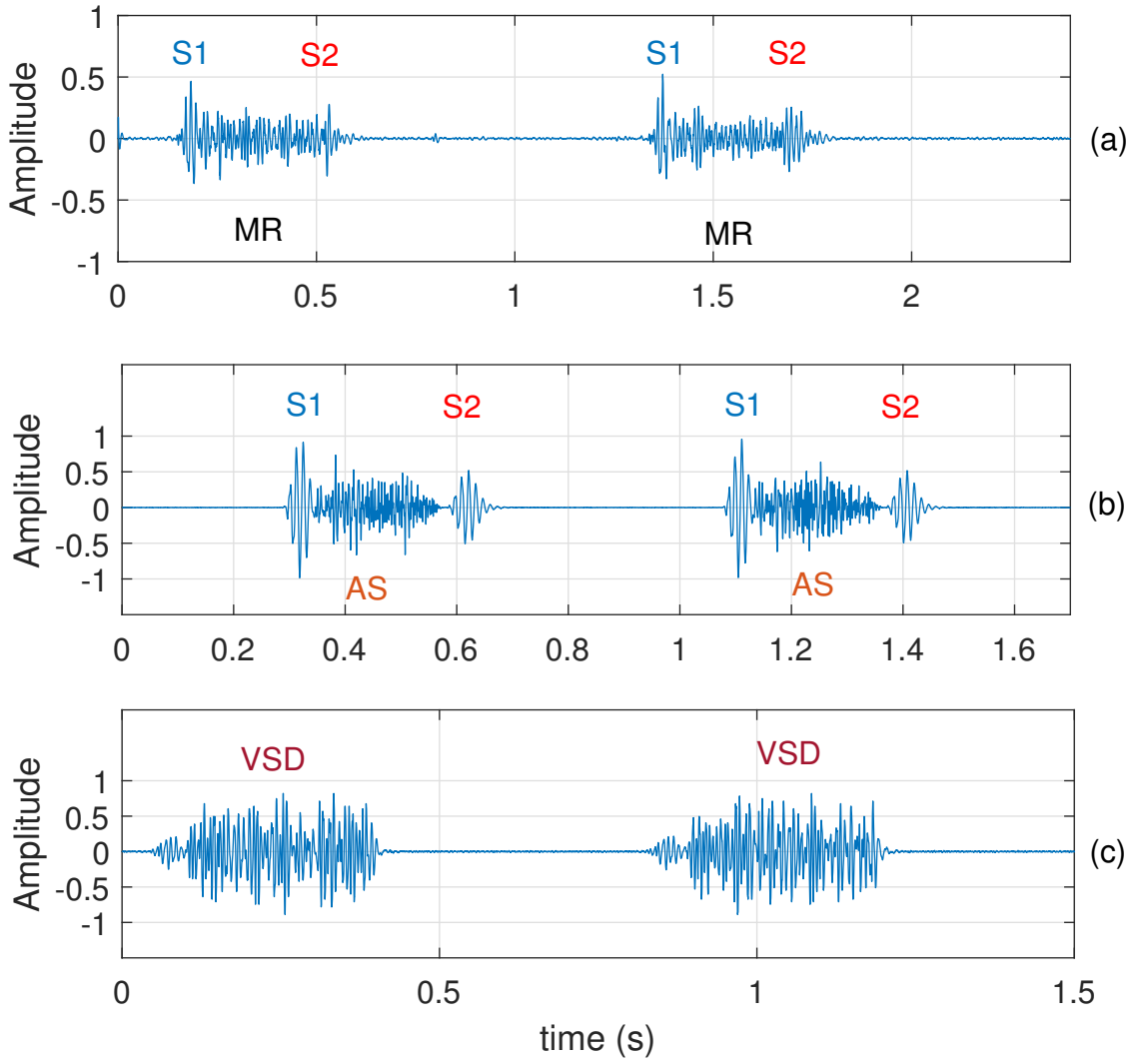

Figure 1.7: Common pathological murmurs: (a) mitral regurgitation (MR), (b) aortic stenosis (AS) and (c) ventricular septal defect (VSD).

Table 1.1: Frequency bandwidth of heart sounds [2].

Heart sound	Frequency bandwidth (Hz)
S3 and S4	15-65
S1 and S2	20-200
Mitral stenosis	40-80
Ejection murmurs	200-400
Regurgitation	250-700

Table 1.2: Symptoms of cardiovascular diseases (CVD) [3].

Pathology	S1	S2	Nature of murmur
AS	normal	single or paradoxically split	mid to late systolic may be soft or absent
MS	louder	normal	diastolic
AR	soft	normal	diastolic & increases with hand grip or squatting
MR	soft	normal or split	systolic and louder
MVP	normal	normal	mid to late systolic

Mitral valve prolapsed (MVP).

With technological advancement, electronic auscultatory devices are available for recording heart sound. The recorded digital signal is called phonocardiography, and the plot is called phonocardiogram or PCG. Phonocardiogram is used for visual inspection of heart sound in clinical diagnosis [25].

1.3 Diagnostic system and diagnostic features

In the mid-1950s, researchers have begun showing interest in the frequency characteristic of the heart sound. It started with an intention to understand the structural and functional aspects of the heart and circulatory system [26]. Rushmer [27], in 1961, hypothesized that the audible heart sounds result from oscillatory over-distention of the myocardium and the valve due to sudden deceleration of blood flow, also known as the 'water hammer' effect. The proposal was backed later by Adolph [28]. The frequency analysis of S1 in myocardial infarction shows that the frequency of heart sounds is directly proportional to the elasticity of the myocardium and valves and inversely proportional to the combined mass of associated heart chambers. The peak frequency of heart sounds may increase or decrease depending

on the heart's health condition. Pioneer suggests that the frequency spectrum of heart sounds may have clinical significance that may be used to develop an automated computer-aided diagnostic system [19, 26, 28, 29].

To develop a CAD system, the analysis of PCG requires identifying the heart sounds and locating their positions. By detecting the S1 and S2 sounds, one can calculate the duration of the cardiac cycle, thus the heart rate. It also helps to identify murmurs in the cardiac cycle as systolic or diastolic murmurs. The murmurs can be further classified into their pathological origins by examining the tone and timbre of the sound. Therefore, developing a robust heart sound segmentation (HSS) algorithm is essential for analyzing heart sound.

A significant number of researchers have dedicated their research efforts to developing a robust HSS algorithm. The existing HSS algorithms can be categorized into the following categories: (i) electrocardiogram (ECG) based method, (ii) morphology-based method, (iii) feature-based method, (iv) probabilistic model-based method, and (v) neural network-based method.

1.3.1 Electrocardiogram-based heart sound detection

An electrocardiogram (ECG) is a record of electrical conduction in the heart responsible for the polarization of heart muscles. This electrical activity can be correlated with valvular events, which in turn, with the PCG signal. As illustrated in Fig. 1.8, the beginning of S1 sound coincides with R-peak, and S2 sound occurs approximately around the end of T-wave of the ECG signal [30–33]. Therefore, the locations of R-peak and T-wave in ECG are considered the gold-standard reference positions for the FHS in the corresponding synchronous PCG. This approach has been adopted in various research works to validate the HSS algorithms. But, using external source signals to identify the heart sound signals is not feasible. Because the outcome of the system wholly depends on the referral signal and its interpretability. Furthermore, ECG is not free of abnormalities, which sometimes is challenging to detect its crucial attributes.

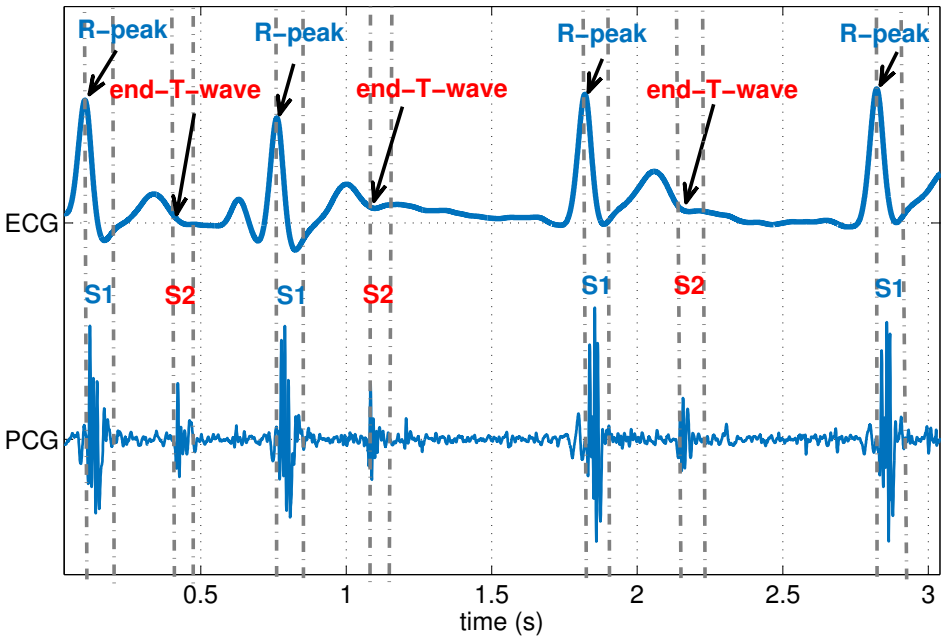

Figure 1.8: PCG and the corresponding ECG.

1.3.2 Morphology-based method

The normal heart sound is perceptually distinguished as loud audible sounds. Here, the intensities of the FHS and their time duration information along the cardiac cycle are crucial clues for deriving the morphological feature. The intensity time graph is calculated as an envelope in signal processing. The envelope of a PCG signal usually produces sharp peaks at the S1 and S2 sounds. The intensity of the peaks is analogous to the loudness level of S1 and S2 sounds. One can identify the heart sound signals and the cardiac intervals by detecting these peaks and comparing the duration difference between the consecutive peaks. Standard methods to extract a signal envelope are by calculating absolute value, signal energy, or derived after Hilbert Transform, or Homomorphic filter [5, 16, 34, 35].

In some cases, Shannon entropy or Shannon energy of the signal is calculated before envelope extraction to uniformly enhance the envelope of FHS [36,37]. Then the second-order derivative of the envelope is calculated to obtain their boundaries. In [38], the teager-kaiser energy operator is applied on the Hilbert envelope of the PCG signal to generate a smooth

envelope surface. Then, the second-order derivative is computed to derive relatively precise boundaries of the FHS. To explore the pertinence of this method on noisy PCG signals, C. D. Papadaniil *et. al.* [39] make use of the kurtosis feature to measure the transition points of FHS based on the bootstrap kurtosis-based criterion. But, using simply the morphology feature to identify and segment the heart sound is not adequate. It is because the envelope extraction methods are sensitive to noise or anomalies in the signal.

1.3.3 Feature-based method

The feature-based method involves tracking the spectrotemporal properties, statistical complexity, correlation of the signal components, or their derivatives along the time axis. A traditional Fourier transform is used in the earlier works to study the spectral evidence of heart sound signal [40, 41]. Alternately, the autoregression method is reported to produce a smoother envelope of signal spectrum [42–45]. The coefficient values may also provide vital information that can be correlated with anomalies of the heart. Therefore, autoregression modeling has been popularly used for the spectral evaluation of PCG signals.

The heart sound, as such, is a short burst sound that exhibits an instantaneous change in amplitude and frequency over time [46, 47]. For inconsistent and highly varying signals, the short-time Fourier transform (STFT) is preferred to evaluate spectrotemporal characteristics. It works by segmenting the signal into smaller segments of equal length by moving a window of fixed frame width across the signal length. From each segmented interval, the Fourier transform is computed. But, examining an STFT over a small signal interval may not track very sensitive sudden changes in time domain [48, 49]. Taking a fixed window also causes resolution problems. If the window size is too short, it cannot resolve the low-frequency components of the signal. On the other hand, if the window size is too long, it becomes difficult to determine the high-frequency components in time.

In the wavelet transform (WT) method, the window size changes depending on the signal frequency composition. A wider window is used for slow varying components and a narrow window for fast varying components. Hence, it gives a better time-frequency resolution of the

signal. In some applications, the discrete wavelet transform (DWT) is used to denoise the PCG signal by selecting feasible details and approximation coefficients that correspond to the frequency band of heart sound and reconstructing the signal back from the coefficients [2,50]. The details and approximation coefficients obtained at different decomposition levels have a non-overlapping frequency band of their own. Therefore, each of these coefficients carries specific spectral and temporal information that may be correlated with heart sound anomalies. There are reported works that use energy, statistical mean and variance, correlation values, or raw coefficient value itself as features to classify the associated pathology [48,51,52].

So far, the wavelet transform has been a potential tool to analyze the subband features of heart sound signals. There is one major drawback of this method. The wavelet used for transformation is predetermined and so it is non-adaptive. The correlation of the selected wavelet with the waveform of heart sounds is essential for successful feature extraction. Once the basic wavelet is selected, it is fixed for all the frequency scales. This may lead to the wrong interpretation of time-frequency variations, mainly observed in detail and approximation coefficients of the lower frequency band. Their characteristics are rather dependent on wavelets and not on the signal from which it is transformed. Therefore, the wavelet-based feature extraction method may not characterize the low-frequency components of heart sounds accurately [35,53]. Among the new signal-dependent spectrotemporal analysis of heart sound, the S-transform, empirical mode decomposition (EMD), and ensemble empirical mode decomposition (EEMD) are also evaluated.

In [36], the S-transform was introduced for the analysis of heart sound. The idea of S-transform originates from both STFT and WT. It has the characteristic of STFT but with a variable window size that depends on frequency components to be analyzed. In S-transform-based HSS, the Shannon energy of the local spectrum after S-transform is used to extract the envelope of the heart sounds. Then an adaptive threshold and peak detection algorithm are used to locate the heart sounds and their borders. The method is evaluated on 80 recordings which have 40 recordings from pathological subjects and yield 96% sensitivity and 95% positive prediction value.

Ari and Saha [53] have introduced EMD to extract features for HSS. It decomposed the signal into oscillatory components called intrinsic mode functions (IMFs) by extracting the energy associated with intrinsic time scales. It preserves the instantaneous frequency of the signal at different levels of IMFs. But the neighboring IMFs tend to have sections of data with the same frequency at different time durations. This mode mixing effect is a major drawback of the EMD method.

This issue is resolved in ensembled empirical mode decomposition (EEMD) by adding white noise to the signal (data) and treating the mean as the final true result [54, 55]. The addition of white noise provides a uniform reference frame in the time-frequency space that makes the different scale signals collate in the proper IMF. In [39], the EEMD and kurtosis features are used to segment heart sounds. Unlike WT, the frequency bandwidth of IMFs obtained after EEMD cannot be determined. Therefore, the selection of IMFs is made by considering the structural frequency of heart sounds while retaining the maximum energy of the signal. It should satisfy the energy-based criterion, i.e., the total energy of the selected IMFs should have 99% of the signal's energy. And, the instantaneous frequency of the selected IMF should be below 150 Hz. Then bootstrap kurtosis-based criterion is used to check if the IMF has either oscillatory or noise-like character. The IMFs that fulfill the said criteria are summed to get the noise-free signal. From this reconstructed signal, the kurtosis feature is calculated again to find the boundaries of the heart sound. Since kurtosis is sensitive to any changes in a signal, it is easily influenced by the presence of noise—any abrupt change consequences in a higher kurtosis value. As a result, a Gaussianity test is performed to remove any unwanted spikes in the kurtosis feature due to noise. The resulting kurtosis feature produces more accurate boundaries of heart sounds. The overall performance is evaluated on 52 recordings consisting of 2608 heart cycles from 11 normal and 16 pathological subjects. It gives an accuracy of 94%.

Tracking instantaneous phase information derived from the analytical signal of a smooth Shannon envelope is evaluated in [56]. It locates the transit point of the phaser waveform to determine the boundaries of FHS. In [5, 16, 57], they use the autocorrelation function of

the PCG envelope to determine the duration of heart sound components and thus locate the S1, and S2 sounds in the signal. Other than amplitude and frequency information, the signal complexity of heart sound may also be a possible feature. In [58], assuming the dynamic system of the human heart, the authors introduced the simplicity measures as the reciprocal of the underlining complexity of FHS derived from eigenvalue-spectrum based on the dynamical systems theory. The simplicity measure and Shannon entropy were used to detect heart sound signals. In [59], the simplicity measure and signal energy at different subbands after wavelet decomposition are used to identify heart sound signals. Then based on the duration information of Cardiac components, the S1 and S2 are further determined and located.

1.3.4 Probabilistic model-based methods

Advanced machine learning tools utilizing probabilistic models have been incorporated into the HSS algorithm. The introduction of the probabilistic model has greatly improved the segmentation performance. Gamero and Watrous [60] have used Mel-frequency cepstral coefficients and the Hidden Markov model (HMM) classifier to segment the heart sound and obtained a reasonable segmentation accuracy. Ricke *et al.* [61] used embedded HMM trained with Mel-frequency cepstral coefficients, regression coefficients, and Shannon energy as system inputs emanated from 2286s recordings achieving an accuracy of 98%. The conventional HMM is based on the first-order Markov process, i.e., the probability of the subsequent outcome depends only on the immediate observation and not on the past observation sequence. Its operation is instantaneous and does not depend on the duration of the state. The probability of transition to the next state remains the same at any time instance. Therefore, the HMM classifier is not ideal for the heart sound segmentation problem. Later, Gill *et al.* [62] have introduced the time duration of heart sounds as an additional feature to model HMM. Using the homomorphic envelope as an input feature and the duration-dependent HMM, also known as the hidden semi-Markov model, as a classifier, the HSS is evaluated on a small dataset recorded from 14 subjects. The performance score obtained is

much better than the conventional HMM classifier. Schmidt *et al.* [16] have used the same HSMM method with sub-band energy of the heart sounds between 25-50 Hz, 50-100 Hz, and 100-150 Hz, and homomorphic envelope as input features to rigorously analyze a noisy dataset. The duration of each heart sound is measured by manual segmenting and listening to each segment of PCG signals. The HSMM is trained on 40 recordings and later tested on 73 recordings of noisy PCG signals. It yields a high positive predictivity value of 98%.

In [5], the logistic regression hidden semi-Markov model (LR-HSMM) based heart sound segmentation algorithm was introduced to evaluate noisy real-time PCG recordings. The synchronous ECG signal is used as a template to validate the location of S1 and S2 sounds in the PCG signal. This information is used only for training and performance evaluation of the LR-HSMM model. The classifier is evaluated using a homomorphic envelope, Hilbert envelope, power spectral density envelope, and wavelet envelope as input features. They have incorporated logistic regression (LR) to calculate the observation probability. The performance is evaluated on a large dataset (approximately 10172 s recordings from 112 subjects) and outperformed the existing results by achieving a F_1 score of 95.63%. The exact duration model and the same extended Viterbi algorithm presented in [5] has been extensively utilized for heart sound segmentation by incorporating different methods of feature extractions or model emission probabilities [63–66]. For optimal estimation of duration distribution, Oliveira *et al.* [64] make use of Gaussian distribution emanated from sojourn time parameters extracted using heuristics approach from the autocorrelation analysis of PCG envelope. In [64], the emission distributions derived using the Gaussian mixture model (GMM) were introduced to fit automatically with the subject. The GMM parameters are determined using the standard expectation maximization (EM) algorithm. This shows the potential of the duration-dependent HSMM for heart sound segmentation problems. With the use of an adequate feature descriptor and emission probability model, there is room to improve the segmentation accuracy.

1.3.5 Neural network-based methods

A more extensive evaluation of LR-HSMM is presented in [67]. It is observed that the fixed sequence tagging of the signal components adopted in the extended Viterbi algorithm makes errors when the heart sounds occur with cardiac arrhythmias, such as large heart rate variation or skipping heartbeats [6]. At present, neural network algorithms such as the convolutional neural network (CNN) and recurrent neural network (RNN) are gaining popularity for HSS research [6, 66, 68–70]. These algorithms have deep learning capability to fit the complex functions, suitable for effective feature classifications, with large feature data [66]. In [69], they proposed a deep neural network (DNN) in combination with the spectral feature derived from mel-frequency cepstral coefficients (MFCC) and K-mean algorithm (to cluster them into refine and discriminative feature groups) for HSS. This method avoids using prior information of state duration and simplifies the training process.

Recurrent neural networks show promising potentials for HSS because of their capability to process sequential input irrespective of the duration and learn the temporal dependency of the sequence. Other neural network structures that learn temporal dependence of input data are the long-short term memory (LSTM) network, gated recurrent neural network (GRNN), and bidirectional recurrent neural network. Their performance for HSS is evaluated elaborately in [6] on large PCG data from the 2016 PhysioNet/CinC Challenge using MFCC, spectrogram, and envelope features. Both bidirectional LSTM (or BiLSTM) and bidirectional GRNN (or BiGRNN) produce comparable results to the state-of-the-art HSMM-based HSS algorithm.

Chen *et al.* [70] have pointed out that extracting predefined feature sets is ambiguous as their performance varies depending on the data used. A particular feature set may give good results in some data but may not produce consistent results when evaluated on the other data. Secondly, feature extraction algorithms may not always extract the desired attributes from all PCG signals and hamper the overall performance. Therefore, in [66, 68] they have attempted to adopt CNN architecture to extract features from PCG signal and used HSMM for temporal classification. Later in [70], the convolutional long short-term memory (CLSTM) was

proposed to use PCG signal as input for the HSS task directly. Here, the convolutional layers play the role of feature extraction, while the LSTM layers execute the sequence recognition task.

1.4 Scope of the work

The use of PCG signal for automated diagnosis of VHD is still in an early phase. In recent year, the envelope features derived from the homomorphic filter, Hilbert transform, Shannon entropy and Shannon energy, and the spectral features such as log-magnitude spectrogram, SBE and MFCC have been explored repeatedly to segment and classify the PCG signal. Studying the temporal relation of heart sound in terms of sound durations, the ratio of the amplitude of heart sounds at different intervals, and their locations in a cardiac cycle have shown promising outcomes that need further investigation for the diagnosis of VHD.

The first step to heart sound analysis is identifying the basic phases of the cardiac cycle (S1, systole, S2 and diastole). In a normal heart sound, these sound segments are capture as the amplitude variations of the PCG envelope. But, any envelope extraction methods is easily affected by noise. In literature, BPF and wavelet transformed based filter are extensively used for PCG denoising. Among nonlinear filters, the total variation filter (TVF) has shown potentials to improve the denoising process and the quality of the PCG envelope. With proper calibration of smoothing parameter by a data-dependent approach, the TVF may become a good noise filter.

The duration-dependent HSMM based HSS is one of the state-of-the-art algorithms in the literature. The study has shown that the single-mode state duration model in the existing HSMM algorithm is not able to accurately segment the PCG signals with irregular heart rate. However, developing a data-dependent duration model may improve the overall performance of the HSMM for HSS.

1.5 Organization of the thesis

The rest of the thesis is organized as follows. Chapter 2 reviews the existing methodologies for preprocessing of PCG signal, popular feature extraction methods, and the HSMM classifier. The motivation of the thesis is included in it. Chapter 3 presents a new adaptive OGS-TVF incorporating sample entropy measure to determine the required degree of smoothing for a PCG signal. It also presents a hybrid dual filtering algorithm combining the BPF with the OGS-TVF for the smart denoising process. In Chapter 4, the logistic function used in the proposed logistic function amplitude moderator (LFAM) is discussed. It discusses the method to obtain optimal parameter values from a given PCG signal to enhance the FHS. Chapter 5 presents a multi-mode diastolic duration model for the HSMM classifier to improve the versatility of the model with the variability of state duration. Chapter 6 presents new frequency bands to calculate the SBE feature. It also introduces the inter-segment correlation coefficients as a discriminative feature for PCG classification. A summary of the overall work is discussed in Chapter 7 with a list of some directions for further research.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

2

Literature Review

Contents

This chapter presents the various methods used for denoising, segmentation, and classification of PCG signal. Developing an intelligent CAD system for analysis of heart sound signals will be of immense advantage for the clinical diagnosis of CVD and to monitor the progress of undergoing treatment. PCG signal carries extensive information on the structural and functional activities of the heart. Any dysfunction of the heart is reflected in the PCG signal as an abnormal sound. Many of these anomalies are quantifiable using advanced signal processing and machine learning techniques. With the growing repository of information derived from the PCG signals, it is becoming more enriching and cognitive for developing an automated diagnostic system.

The reviews of the prevailing popular algorithms for the analysis of PCG signals are discussed in details in the following sections. In Section 2.1, the PCG database used for the evaluation of the thesis work is discussed. Section 2.2 presents the denoising process using wavelet decomposition and total variation filter. Both the denoising methods were considered to produce better noise reduction of PCG signals. Section 2.3 discusses some popular feature extraction methods for the analysis of heart sound signals. In Section 2.4, HSMM based heart sound segmentation algorithm is reviewed. The performance matrices that will be used for evaluating the thesis work is discussed in Section 2.5. The motivation of this thesis is presented in Section 2.6.

2.1 Database

In this thesis, two open-access databases: one by PhysioNet/Computing in Cardiology Challenge 2016 and the other by the PASCAL classifying heart sounds challenge 2011 (CHSC2011) are evaluated. They are briefly discussed as follows.

2.1.1 PCG Data

The heart sound database by PhysioNet/CinC Challenge 2016 is assembled from different independent sources. Description of the databases is mention in Ref. [18]. The URL for the

location of PCG files is [71]. They are available as the training set (a through f), in total 3,153 heart sound recordings from 764 subjects/patients [4].

An example dataset used in [5], which is downloadable from [72], available as training set-a by MIT [4, 18] is briefly discussed as follows. The Massachusetts Institute of Technology heart sound database (MITHSDB) is contributed by Syed *et al* [73]. The PCG signals are recorded simultaneously with an electrocardiogram (ECG) using a Meditron electronic stethoscope. The sample rate of the recordings is 44.1 kHz with 16-bit quantization. There are a total of 409 recordings from 121 subjects. The length of the recordings ranges from 9 s to 37 s. The database comprises 117 recordings from 37 healthy subjects, 134 recordings from 37 patients with symptoms of mitral valve prolapse (MVP), 118 recordings from 34 patients showing innocents or benign murmurs (Benign), 17 recordings from patients with aortic diseases (AD), and 23 recordings from 7 patients that have other miscellaneous pathological conditions (MPC). The recordings are performed by either visiting the patients at their homes or in the hospital. As a result, many of the recordings are corrupted with various noise sources from the surrounding. Other noise sources include respiratory noise, intestinal sounds, and artifacts from stethoscope motions. This particular dataset provides corresponding ECG signals for validating the heart sound components. Therefore, it is used in this thesis to evaluate different proposed modules involved in developing the HSMM-based HSS algorithms. That includes the denoising process in Chapter 3, the envelope enhancement methods in Chapter 4 and the proposed duration models for the HSMM-based HSS algorithm in Chapter 5.

The above-discussed dataset lag details information to categorize heart sound murmur from other pathological PCG recordings. Therefore, to evaluate different diagnostic features for the classification of heart sound into categories: normal, noisy, and murmur, in Chapter 6, a separate dataset available in the 'classifying heart sounds challenge' sponsored by PASCAL [74] is used. The database by the PASCAL/CHSC2011 has been gathered from two sources: Dataset A is from the general public via the iStethoscope Pro iPhone app and Dataset B is from a clinical trial in hospitals using the digital stethoscope DigiScope [74]. But details of the subject or pathology associated with the signals are not specified. Dataset

Table 2.1: Profiles of PCG database.

Database	Data Set	Category	Recording state	# Recordings	Simultaneous signals	Sampling rate	Sensor
PhysioNet/ CinC Challenge 2016	a (MITHSDB)	Normal MVP Benign AD MPC	Uncontrolled environment (in-home visits or in hospital)	117 134 118 17 23	one PCG one ECG	44.1 kHz	Meditron electronic stethoscope
PASCAL/ CHSC2011	A	Normal Murmur Extra heart-sound Artifact	Uncontrolled environment from general public	31 34 19 40	one PCG	44.1 kHz	iStethoscope Pro iPhone app
	B	Normal Murmur Extrasystole	Clinical trial in hospitals	200 66 46	one PCG	4000 Hz	Digital stethoscope DigiScope

A is categorised into four classes, namely: normal, murmur, extra heart sound and artifact. Category wise, there are 31 recordings in the normal, 34 recordings in the murmur, 19 recordings in the extra heart sound and 40 recordings in the artifact category. Dataset B is subcategorised into normal, murmur and extrasystole. There are 200 recordings in the normal, 66 recordings in the murmur and 46 recordings in the extrasystole category. The recordings might have been collected from children and adults in both calm and excited states. The heart rates may vary from 40 to 140 bpm [75, 76]. Some recordings are corrupted with background noise (from traffic to radio), physiological sound and artifactes from recording devices (movement of microphone, and connecting wires). In the murmur category, the types of murmurs are not specified. However, it contains a wide range of murmurs having temporal locations in systole or diastole intervals and with varying degrees of intensities. Some of the recordings are noisy due to background noise or distortion. The extra heart sound category of Dataset A is the collection of recordings containing additional sounds in the normal heart sound cycle. It may include split sounds or gallop sounds. The artifact category (of Dataset A) is a collection of a wide range of noise including feedback squeals and echoes, speech, music and background noise. There are no discernible heart sounds and purely meant to indicate possible noise signals. The extrasystole category (Dataset B) is an occasional occurrence

in which the heart sound is out of rhythm. There may be extra or skipped heart beats. The detailed profiles of the PCG databases discussed above are tabulated in Table 2.1.

2.1.2 Noise database

Various ambient noise is added into the original signal to simulate noisy PCG recordings from a noisy hospital-like environment. The noise signals are downloaded from the open web database Freesound (`https://freesound.org`). They are broadly categorized as (a) ventilation/air conditioner (AC) noise, (b) ambulance noise, and (c) hospital ambient noise. There are nine recordings of AC noise type which may be affected by faulty leakage, crickets, broken fan, angered sound, and additional street and footstep noise. For ambulance noise, there are 13 recordings consisting of vehicle sounds, sirens, people talking, rain, and traffic sounds. There are 12 hospital ambient noise recordings consisting of crowd (nurse, doctor, patient) noise, coughing, door opening or closing, footsteps, humming sound of various machines, the beeping of heart machine recorded at hospital corridor, emergency room, and ICU noise in it. There is no discernible heart sounds in the noise signals. In the following chapters, these noises will be known as real-time ambient noise (RTAN).

2.2 Denoising process

PCG signals are not always recorded in an ideal environment. As discussed in the earlier section, a PCG signal is prone to noise and artifacts intruded through external influences. The presence of noise may affect the way the PCG signal is interpreted. So there is a need for preprocessing that can eliminate these noises without altering the original PCG signal [3]. The popular methods of PCG denoising are based on wavelet transform (WT), singular value decomposition (SVD), empirical mode decomposition, bandpass filter (BPF), adaptive filtering, etc. Some applications make use of combining these methods for effective denoising of PCG signals.

2.2.1 Wavelet transform (WT)

The wavelet transform was developed as an improvement over STFT to obtain a higher resolution of time and frequency information simultaneously by adopting appropriate window size for different frequency bands. It makes use of a wavelet, which is either compressed or stretched across the set frame size to generate coefficients corresponding to different frequency compositions. Narrowing the wavelet reciprocate the high-frequency components of the signal as detail coefficients, while stretching will extract the lower frequency components in terms of approximation coefficients. The detail coefficients d_m and the approximation coefficients a_m of m level discrete wavelet transform are calculated from scaled and shifted version of wavelet function ψ_m and scaling function ϕ_m [77] as:

$$d_m(k) = \frac{1}{\sqrt{M}} \sum_n x(n)\psi_m(\frac{n-k}{2^m}) \tag{2.1}$$

$$a_m(k) = \frac{1}{\sqrt{M}} \sum_n x(n)\phi_m(\frac{n-k}{2^m}) \tag{2.2}$$

These coefficients can be calculated efficiently by fast discrete wavelet transform algorithm using analysis filters coefficients $\mathbf{h}$ and $\mathbf{g}$ (high pass and low pass, respectively):

$$d_m(k) = h(n) * a_{m-1}(n)|_{n=2\times k, k\geqslant 0} \tag{2.3}$$

$$a_m(k) = g(n) * a_{m-1}(n)|_{n=2\times k, k\geqslant 0} \tag{2.4}$$

The signal can be reconstructed from the detail and approximation coefficients as follows [78]:

$$x(n) = \sum_k a_M(k)\tilde{g}(k-2n) + \sum_{m=1}^{M}\sum_k d_m(k)\tilde{h}(k-2n) \tag{2.5}$$

where $\tilde{g}$ and $\tilde{h}$ are the synthesis low-pass and high-pass filters, respectively. The analysis and synthesis filters together form the quadrature mirror filters which follow alternating flip and alternating sign change of filter coefficients [78, 79].

WT has a wide range of applications for PCG analysis. The level-wise decomposition of the signal into a layer of smaller sub-bands helps analyze the PCG signal effectively.

A standard method of WT-based denoising of PCG signal is by selecting the details and approximations which correspond to the frequency band of heart sound and using them to reconstruct the signal back [50, 78, 80]. Jain *et al.* [78] used "Coif-5" as the mother wavelet of WT and adopted an adaptive threshold estimation method based on the non-linear mid function to discard the smaller detail coefficients from the decomposition. The reconstructed signal from the remaining coefficients improved the elimination of out-of-band noise and preserved most of the heart sound attributes intact.

A hybrid of wavelet packet transform (WPT) and singular value decomposition (SVD) is an improvement over the previous method [81]. The algorithm involves selecting a descriptive node of WPT that satisfies the criterion of subband and mutual information measurement conditioning to the original heart sound. The selected nodal coefficients were further processed using SVD to remove the residual noise in the WPT and refine the resulting denoised PCG signal.

2.2.2 Total variation filter (TVF)

The TVF is a nonlinear data-dependent low-pass filter capable of restoring piecewise-constant signals. It works based on the Tikhonov regularization smoothing strategy employing L_1 norms of derivatives or total variation (TV). It was first introduced by Rudin, Osher, and Fatemi [82] to remove oscillatory noise components without distorting edges of an image. This model has wide application in image restoration, deconvolution, interpolation, and compressed sensing [83, 84]. In this section, we briefly discussed an algorithm for TVF and its applications in the heart sound signal.

Given an N-point noisy signal $\mathbf{x} = \mathbf{y} +$ noise, the TVF estimates the expected signal $\bar{\mathbf{y}}$ retaining its total variation, i.e. $\text{TV}(\mathbf{y}) = \|\mathbf{D}\mathbf{y}\|_1 = \sum_{n=2}^{N} |y(n) - y(n-1)|$. It is solved as an unconstrained optimization problem with an objective function defined under:

$$\arg\min_{\mathbf{y}} \left(F(\mathbf{y}) = \left[\frac{1}{2} \sum_{n=1}^{N} (x(n) - y(n))^2 + \lambda \|\mathbf{D}\mathbf{y}\|_1 \right] \right) \tag{2.6}$$

where $\mathbf{D}$ is a derivative matrix of size $(N-1) \times N$

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & -1 & 1 \end{bmatrix}$$

and $\lambda > 0$ is a regularization tuning parameter that controls the weight of the total variation. Increasing λ will increase the degree of smoothing. There is no explicit solution to Eq. (2.6). One of the fast algorithm adopted in our work is derived using the majorization-minimization (MM) optimization [83] (available with a Matlab implementation [85]).

The MM is a procedure to simplify a difficult minimization problem by replacing it with a sequence of simpler ones. Instead of minimizing a function $F(\mathbf{y})$ directly, the MM produces a sequence $\bar{\mathbf{y}}_i$ at each iteration by solving an equivalent function $G_{i-1}(\mathbf{y})$ that converges towards the minimizer of $F(\mathbf{y})$.

$$\bar{\mathbf{y}}_{i+1} = \arg\min_{\mathbf{y}} G_i(\mathbf{y}) \tag{2.7}$$

where i is the iteration index; $G_i(\mathbf{y})$ is a convex majorizer of $F(\mathbf{y})$ such that $G_i(\mathbf{y}) \geq F(\mathbf{y})\ \forall \mathbf{y}$ and coincides with $F(\mathbf{y})$ at $\bar{\mathbf{y}}_i$ (i.e. $G_i(\mathbf{y}_i) = F(\mathbf{y}_i)$). The function $G_i(\mathbf{y})$ should be easier to solve than the parent function $F(\mathbf{y})$.

The minimization of the objective function $F(\mathbf{y})$ in Eq. (2.6) is difficult due to the fact that the L_1 norm is not differentiable. Therefore, an equivalent quadratic polynomial $g(y) = \frac{1}{2|y_i|}y^2 + \frac{1}{2}|y_i|$ that majorizes $f(y) = |y|$ and agrees with $f(y)$ at $y = y_i$ is introduced. It satisfy

$$\frac{1}{2\,|y_i|}y^2 + \frac{1}{2}\,|y_i| \geqslant |y|\,, \qquad \forall y \in \mathbb{R} \tag{2.8}$$

Summation of the above expression over the whole signal length can be compactly expressed as

$$\frac{1}{2}\mathbf{y}^T \Lambda_i \mathbf{y} + \frac{1}{2}\,\|\mathbf{y}_i\|_1 \geq \|\mathbf{y}\|_1 \tag{2.9}$$

where Λ_i is a diagonal matrix

$$\Lambda_i = \begin{bmatrix} |y_i(1)|_1^{-1} & & & \\ & |y_i(2)|_1^{-1} & & \\ & & \ddots & \\ & & & |y_i(N)|_1^{-1} \end{bmatrix}$$

The notation, $|*|$ is an absolute value applied element wise.

Introducing (2.9) in the expression (2.6) of the cost function $F(\mathbf{x})$, the majorizer function $G_i(\mathbf{x})$ can be obtained.

$$\underbrace{\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{D}^{\mathbf{T}}\Lambda_i\mathbf{D}\mathbf{y} + \lambda\frac{1}{2}\|\mathbf{D}\bar{\mathbf{y}}_i\|_1}_{G_i(\mathbf{y})} \geq \underbrace{\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{y}\|_1}_{F(\mathbf{y})}, \quad \Lambda_i = \text{diag}(|\mathbf{D}\bar{\mathbf{y}}_i|)^{-1}) \tag{2.10}$$

which is in the form of $G_i(\mathbf{y}) \geq F(\mathbf{y})$ and satisfies $G_i(\mathbf{y}_i) = F(\mathbf{y}_i)$ by design. Then, the convergent sequence $\bar{\mathbf{y}}_i$ is obtained by minimizing $G_i(\mathbf{y})$.

$$\bar{\mathbf{y}}_{i+1} = \arg\min_{\mathbf{y}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{D}^{\mathbf{T}}\Lambda_i\mathbf{D}\mathbf{y} + \lambda\frac{1}{2}\|\mathbf{D}\bar{\mathbf{y}}_i\|_1 \tag{2.11}$$

The solution to Eq. (2.11) is given by

$$\bar{\mathbf{y}}_{i+1} = \left(\mathbf{I} + \lambda\mathbf{D}^{\mathbf{T}}\Lambda_i\mathbf{D}\right)^{-1}\mathbf{x} \tag{2.12}$$

If the sparsity of $\mathbf{D}\bar{\mathbf{y}}_i$ increases, some entries of the diagonal matrix Λ_i may become infinity causing numerical inaccuracy in the update equation Eq. (2.12). This issue may be resolved by rewriting the expression using the matrix inverse lemma [83, 86].

$$\left(\mathbf{I} + \lambda\mathbf{D}^{\mathbf{T}}\Lambda_i^{-1}\mathbf{D}\right)^{-1} = \mathbf{I} - \mathbf{D}^{\mathbf{T}}\left(\frac{1}{\lambda}\mathbf{\Lambda}_i^{-1} + \mathbf{D}\mathbf{D}^{\mathbf{T}}\right)^{-1}\mathbf{D} \tag{2.13}$$

The expression Eq. (2.12) becomes:

$$\bar{\mathbf{y}}_{i+1} = \mathbf{x} - \mathbf{D}^{\mathbf{T}}\left(\frac{1}{\lambda}\mathbf{\Lambda}_i^{-1} + \mathbf{D}\mathbf{D}^{\mathbf{T}}\right)^{-1}\mathbf{D}\mathbf{x} \tag{2.14}$$

or:

$$\bar{\mathbf{y}}_{i+1} = \mathbf{x} - \mathbf{D}^{\mathbf{T}} \left(\frac{1}{\lambda} \text{diag}(|\mathbf{D}\bar{\mathbf{y}}_i|) + \mathbf{D}\mathbf{D}^{\mathbf{T}} \right)^{-1} \mathbf{D}\mathbf{x}$$

For more details, please refer [83, 86, 87].

In general, any heart sound analysis system uses a simple digital low-pass filter (LPF) operating at a cut-off frequency of 150 Hz. It is a feasible choice considering the fact that most of the signal energy of FHS is below 150 Hz [19]. But LTI filters are suitable when the signal and noise have a distinct and separable frequency band. If a signal is not isolated to a specific frequency band and has piecewise discontinuities, the filter outcome may suffer from residual frequencies due to spectral leakage. PCG signal is one such example that has silent systole and diastole interval separating the heart sound signals. For such signals, sparsity-based denoising is appropriate.

In [56], the TVF is incorporated as a primary denoising module in the heart sound activity detection framework. The smoothing technique not only removes high-frequency components but also suppresses low amplitude in-band noise in the discontinuing intervals of PCG. But this method has one major challenge that needs addressing. That is fixing the value of λ. There is no conclusive method to decide the degree of smoothing required for different levels of noise. Most articles have reported using a hard threshold value based on experimental results. Excessive smoothing may also cause clipping of lower amplitude signal waveform, prominently seen when the noise intensity is much higher than the actual heart sound signals. Therefore there is a need to establish the regularization parameter λ value from the nature and complexity of the signal and the noise level.

2.2.3 Overlapping group sparsity denoising

The use of total variation as a penalty function in sparse signal processing tends to produce stair-case artifacts. The resulting signal may have sharper edges or clipping, which is not natural for sinusoidal signals. To improve the non-linear denoising method, a more generic and versatile approach focusing on structured sparsity has been adopted [88]. The method considers the possibility of several large values of derivatives to arise adjacent to other

large values exhibiting group sparsity. The grouping behavior is observed in many naturally occurring signals such as the PCG signal where large derivative values are localized around the S1 and S2 sounds. In order to comprehend the concept, the penalty function is derived based on overlapping group sparsity (OGS) [1, 88, 89]. A K-point vector of the sequence $\mathbf{x}$ representing a group is denoted by:

$$\mathbf{x}_{n,K} = [x(n), ..., x(n+K-1)]$$

The OGS penalty function is defined as:

$$\varphi(\mathbf{x}) = \sum_{n=1}^{N-K+1} \vartheta\left(\mathbf{x}_{n,K}\right)$$

with the objective function expressed as

$$\arg\min_{\mathbf{y}} \left\{ F(\mathbf{y}) = \frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|_2^2 + \lambda\varphi\left(\mathbf{D}\mathbf{y}\right) \right\} \tag{2.15}$$

where ϑ can be L1-norm, L2-norm, logarithmic, or any concave function [1].

If $K = 1$ and $\vartheta(*)$ takes L1-norm, the problem definition in Eq. (2.15) becomes the standard TVF problem. If $K > 1$ and $\vartheta(*)$ operates as L2-norm, the problem definition is OGS-TVF. The solution to this optimization problem follows the same MM iterative process of TVF given by Eq. (2.14) with slight changes in the penalty function φ and entries of the diagonal matrix Λ_i [1, 88].

$$\varphi(\mathbf{x}) = \sum_{n=1}^{N-K+1} \left[\sum_{k=0}^{K-1} |x(n-k)|^2 \right]^{1/2} \tag{2.16}$$

$$[\Lambda_i(\mathbf{y}_i)]_{n,n} = \lambda \sum_{j=0}^{K-1} \left[\sum_{k=0}^{K-1} |y_i(n-j+k)|^2 \right]^{-1/2} \tag{2.17}$$

The method alleviates the staircase artifacts in the denoised signal as an improvement over the standard TVF. The algorithm of OGS-TVF is summarized in Algorithm 1.

Algorithm 1 OGS-Total variation denoising algotirthm.

Initialize

$\bar{\mathbf{y}}_1 = \mathbf{x}$

$i = 1$

repeat

$[\Lambda_i(\bar{\mathbf{y}}_i)]_{n,n} = \lambda \sum_{j=0}^{K-1} \vartheta \left([\mathbf{D}\bar{\mathbf{y}}_i]_{n-j,K} \right)^{-1}$

$\mathbf{F} = \frac{2}{\lambda}\Lambda_i^{-1} + \mathbf{D}\mathbf{D}^{\mathbf{T}}$ ($\mathbf{F}$ is tridiagonal)

$\bar{\mathbf{y}}_{i+1} = \mathbf{x} - \mathbf{D}^{\mathbf{T}}(\mathbf{F} \backslash \mathbf{D}\mathbf{x})$ (fast solver)

$i = i + 1$

until convergence or satisfy a condition

2.2.3.1 Estimating adaptive regularization parameter (λ)

The OGS regularization may also be realized by estimating the likelihood of a denoised signal with prior knowledge of the original signal (clean) and the noise. The denoising problem is solved using the maximum a-posteriori (MAP) estimation of the original signal based on the hierarchical Bayesian inference [1]. Assuming that the noisy signal $\mathbf{x} = \mathbf{y} +$ noise has zero mean with known variance ϱ^2 value, and it is uniformly distributed across the signal, the likelihood of $\mathbf{x}$ may be defined as:

$$P\left(\mathbf{x}|\mathbf{y}, \varrho^2\mathbf{I}\right) \propto e^{\frac{-\|\mathbf{x}-\mathbf{y}\|_2^2}{2\varrho^2}} \tag{2.18}$$

Taking γ as the hyperparameter involved in the OGS penalty function $\varphi(\mathbf{D}\mathbf{y})$, the prior knowledge is formulated as:

$$p\left(\mathbf{y}|\gamma\right) \propto \frac{1}{\gamma^{-\theta N}} e^{-\gamma\varphi(\mathbf{D}\mathbf{y})} \tag{2.19}$$

where $\gamma^{-\theta N}$ is a normalization factor and $\theta \in (0, 1)$ is constant. Assigning the hyperparameter $\gamma(> 0)$ in terms of the Gamma prior as in [90], the likelihood value is obtained below:

$$p\left(\gamma|\grave{\alpha}, \grave{\beta}\right) \propto \gamma^{\grave{\alpha}-1} e^{-\grave{\beta}\gamma} \tag{2.20}$$

where $\grave{\alpha}$ and $\grave{\beta}$ are the shape and scale parameters, respectively. By Bayes' theorem, the prior distribution of the original signal $\mathbf{y}$ is obtained by the joint probability distribution

$P(\mathbf{y}) = \int_0^{+\infty} p\left(\mathbf{y}|\gamma\right) p\left(\gamma|\grave{\alpha}, \grave{\beta}\right) d\gamma$ which can be approximately represented as [1]:

$$P(\mathbf{y}) = \left(\varphi(\mathbf{Dx}) + \grave{\beta}\right)^{-(\theta N + \grave{\alpha})} \tag{2.21}$$

Using Eq. (2.18) and Eq. (2.21), the posterior distribution of $\mathbf{y}$ is calculated as:

$$\begin{aligned} P(\mathbf{y}|\mathbf{x}) \quad &= P(\mathbf{x}|\mathbf{y}, \varrho^{\mathbf{2}}\mathbf{I}) P(\mathbf{y}) \\ &\propto e^{-\frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{2\varrho^2}} \left(\varphi(\mathbf{Dx}) + \grave{\beta}\right)^{-(\theta N + \grave{\alpha})} \end{aligned} \tag{2.22}$$

Detailed discussion on how to obtain the expressions are presented in [90] and the references therein. The MAP estimation of the original signal from the log-likelihood of Eq. (2.22) is presented as:

$$\mathbf{y}_{MAP} = \arg\min_{\mathbf{y}} \left(\underbrace{\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \varrho^2(\theta N + \grave{\alpha}) \log(\varphi(\mathbf{Dy}) + \grave{\beta})}_{L(\mathbf{y})} \right). \tag{2.23}$$

This optimization problem is solved using the MM algorithm by constructing a majorizer function of $L(\mathbf{y})$ using the following log inequality [90].

$$\log(v) \leq \log(v_0) + \frac{v - v_0}{v_0} \quad \forall v, v_0 \in \mathbb{R}_+ \tag{2.24}$$

That gives,

$$\log(\varphi(\mathbf{Dy} + \grave{\beta})) \leq \frac{\varphi(\mathbf{Dy})}{\varphi(\mathbf{Dv}) + \grave{\beta}} + C \tag{2.25}$$

where C is independent of $\mathbf{y}$. Using Eq. (2.25), the mazorizer of $L(\mathbf{y})$ can be formulated as:

$$Q(\mathbf{y}, \mathbf{v}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \rho(\mathbf{v})\varphi(\mathbf{Dy}) + C_1 \tag{2.26}$$

where C_1 is independent of $\mathbf{y}$ and $\rho(\mathbf{v}) = \frac{\varrho^2(\theta N + \grave{\alpha})}{\varphi(\mathbf{Dv} + \grave{\beta})}$ Eq. (2.26) satisfies the condition for being a convex majorizer $Q(\mathbf{y}, \mathbf{v}) \geq L(\mathbf{y})$, $\forall \mathbf{v}$ and concides at $\mathbf{v} = \mathbf{y}$, i.e. $Q(\mathbf{y}, \mathbf{y}) = F(\mathbf{y})$. On comparing the expression with Eq. (2.15), $\rho(\mathbf{v})$ may be viewed as the estimation of λ at $\mathbf{v}$.

The iterative MAP estimation of $\mathbf{y}$ using the MM algorithm is represented by:

$$\begin{aligned}\bar{\mathbf{y}}_{i+1} &= \arg\min_{\mathbf{y}} \left(Q(\mathbf{y}, \bar{\mathbf{y}}_i)\right) \\ &\triangleq \mathbf{x} - \mathbf{D}^{\mathbf{T}} \left(\tfrac{1}{\lambda_i}\mathbf{\Lambda}_i^{-1} + \mathbf{D}\mathbf{D}^{\mathbf{T}}\right)^{-1} \mathbf{D}\mathbf{x}\end{aligned} \tag{2.27}$$

and

$$\lambda_{i+1} = \rho(\bar{\mathbf{y}}_{i+1}) \tag{2.28}$$

It follows the same algorithm illustrated in Algorithm 1. The iteration process continues until the denoised signal $\bar{\mathbf{y}}_i$ resembles the original signal. If the denoised signal exactly equals the original signal, the residual signal $\mathbf{x} - \bar{\mathbf{y}}_i$ will be the noise signal, and their variance will bear the same value.

$$\varrho^2_{\text{noise}} = \frac{1}{N} \|\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.29}$$

which implies

$$\|\mathbf{x} - \mathbf{y}_i\|_2^2 \leq N\varrho^2_{\text{noise}} \tag{2.30}$$

Eq. (2.30) is used as a stop condition of the OGS-TVF. In [1], the OGS-TDF formulated with adaptive MAP estimation of regularization parameter was evaluated for heart sound denoising. In their experiment, the variance of the noise is calculated by the median absolute deviation (MAD) rule [91]. The remaining parameters $\grave{\alpha}$, $\grave{\beta}$, θ and group size K values were pre-set at 1, 50, 0.8 and 20, respectively. The denoising performance showed significant improvement over the conventional wavelet-based method.

2.3 Feature extraction

The morphology and spectral property are two principal domains of feature for any heart sound analysis. In the morphological feature, the envelope of PCG signal is explored to extract temporal information such as the possible location of FHS as envelope peaks and the approximate duration of the cardiac cycle, S1-S2 or Systolic interval and S2-S1 or diastolic interval. Intuitively, this information are used by physician and experts for analysis of the heart sound signals. Many researchers have also exploited the evidence that the fundamental

heart sounds have the majority of the signal energy below 150 Hz with a peak at 50 Hz. The features for heart sound segmentation problem are defined to closely meet the above description.

2.3.1 Envelop extraction

The envelope helps to visualize the distribution of signal intensities at different time instances. Based on the mechanism of producing a normal heart sound, the S1 and S2 sounds are only audible sounds separated by silent intervals. Therefore, a PCG envelope reproduces the temporal characteristic of the heart sound signal that can be used to interpret the intensity of FHS. Most envelope extraction method for HSS are derived from either Shannon Entropy, Shannon Energy, Hilbert Transform, or Homomorphic envelope [5, 16, 34, 35]. There are other primitive methods to calculate envelope based on energy (squaring) or absolute value [34]. In an energy-based envelope, the weaker signals that have smaller intensity ration are suppressed when compared to stronger intensity signals. On the other hand, the absolute value gives equal weight to all the intensities. The expressions for different envelogram methods are as under:

$$\text{energy} \quad = \quad x^2(t) \tag{2.31}$$

$$\text{absolute value} \quad = \quad |x(t)| \tag{2.32}$$

Hilbert envelope ($\mathrm{Hilbert\_E}$): Envelope from analytic signal is a very common envelogram method [5, 50, 92]. The analytic signal is obtained after Hilbert transform of a signal which is given by:

$$y(t) = x(t) + j\hat{x}(t) \tag{2.33}$$

where, $x(t)$ is the original signal and its Hilbert transform $\hat{x}(t)$ is defined as:

$$\hat{x}(t) = \frac{1}{\pi t} * x(t) \tag{2.34}$$

The Hilbert envelope is calculated from the analytic signal using the following equations:

$$\text{Hilbert\_E} = \sqrt{x^2(t) + \hat{x}^2(t)} \tag{2.35}$$

Homomorphic envelope (Homo_E): Another popular envelope extraction method is using homomorphic filter [35,62]. The advantage of homomorphic envelope is its scalable smoothing process, which can be easily tuned. It extracts the slow varying component $e(t)$ of a signal that envelops the fast oscillating components $o(t)$.

$$x(t) \quad = e(t) \cdot o(t) \tag{2.36}$$

Natural logarithm operator is used to convert multiplication to summation of the two components. This additive nature makes the high-frequency component appear as a rapid variation in time.

$$\ln x(t) \quad = \quad \ln e(t) \; + \; \ln o(t) \tag{2.37}$$

Applying a simple low pass filter (LPF) of desired cut-off frequency will remove the unwanted frequency components. The signal is retrieved by exponentiating the filter output.

$$\text{Homo\_E} \quad = \quad \exp\left[\text{LPF}\left[\ln\ x(t)\right]\right] \quad \approx \quad e(t) \tag{2.38}$$

Envelope moderation methods: For the uniformity and easy detection of FHS, it is preferred that the envelope peaks corresponding to S1 and S2 should maintain more-or-less identical intensity levels. To realise this, Shannon entropy and Shannon energy are calculated before envelope extraction. Shannon entropy and Shannon energy emphasize the medium intensity signals by using logarithmic weight. They are calculated as:

$$\text{Shannon energy} = -x^2(t)\log(x^2(t)) \tag{2.39}$$

$$\text{Shannon entropy} = -|x(t)|\log|x(t)| \tag{2.40}$$

2.3.2 Frequency domain features:

The popular frequency domain features for HSS are power spectral density envelope and wavelet envelope [5, 6]. There will be no denoising process before the feature extraction process.

Power spectral density (PSD) envelope: From the spectral analysis of heart sound, it is obvious that the spectral peak occurs approximately at 50 Hz [5]. This is true for a healthy heart sound. However, in the case of a pathological condition such as Myocardial Infarction, cardiomyopathy, and valvular defects, the spectral peak of the S1 and S2 sounds may deviate depending on the elastic property or thickness of myocardium, and the pressure built up in the heart chambers [28]. The exact peak frequencies associated with different pathological cases are not fully established. Therefore, it is better to include spectral peak within the frequency range of 20 Hz to 150 Hz, because, the major signal energy of FHS is below 150 Hz [19]. Thus, the PSD envelope is calculated as the frequency of the spectral peak within the permissible range. The Hamming window of size 50 ms is used for short-term-Fourier-transform of the PCG [5].

Wavelet envelope: Wavelet analysis has been extensively used for the analysis of heart sound signals. But choosing the wavelet family suitable for the purpose is still a difficult decision. In the related work [5], wavelet families inclusive of Haar, Daubechies, symlet, Coiflet, biorthogonal, and reverse biorthogonal wavelets were used to decompose the signal into 1-5 levels. The detail coefficients and the corresponding levels that fit the frequency band of the normal heart sound were selected. The envelope value is the sum of the absolute

values of detail coefficients. To check how well the wavelet discriminates the FHS, the ratio of the sum of the absolute values of selected coefficients corresponding to S1 and S2 sounds against other intervals and across all recordings is measured. The wavelet yielding the highest ratio value is considered the best choice.

2.4 Hidden semi-Markov model (HSMM)

Hidden markov model (HMM) is a doubly embedded stochastic framework used to characterize the sequential data into discrete hidden states by analyzing the observed features. The model is governed by the first-order Markov assumption [93]. For 'N' number of individual states $S = \{S_1, S_2, S_3, ..., S_N\}$ and given observation sequence $O = \{O_1, O_2, O_3, ..., O_T\}$, where O_t is an instantaneous observation generated at any hidden state q_t, the HMM is defined as

$$\hat{\Lambda} = \{A, B, \pi\} \tag{2.41}$$

where $A = \{a_{ij}\}$ is the transition matrix defined as the probability of transition from state 'S_i' to state 'S_j'.

i.e.

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i) \tag{2.42}$$

$B = \{b_j(O_t)\}$ is the probability of observation 'O_t' being generated at state 'S_j'.

i.e.

$$b_j(O_t) = P(O_t | q_t = S_j) \tag{2.43}$$

and $\pi = \{\pi_i\}$ is the initial state probability being in state 'i'. i.e.

$$\pi_i = P(q_1 = S_i) \tag{2.44}$$

The HMM model is successfully used in many research areas including medical applications to classify large dataset which are highly variable within the same class. In heart sound analysis, the model is used to compute the single best state sequence $\hat{Q}$ which maximizes

the expectation of the given observation sequence O from a given model $\hat{\Lambda}$.

$$\hat{Q} = \arg\max_{Q} \; P(Q|O, \hat{\Lambda}) \tag{2.45}$$

The traditional method of solving equation (2.45) is to calculate likelihood for all possible combinations of state sequences Q and select the sequence which holds maximum probability. This approach is computationally costly. Therefore, a dynamic programming method called the Viterbi algorithm is used to find the optimal state sequence. It involves estimation of the instantaneous probability $\delta_t(i)$ that yields the best score along the path defined by the first t observations and ends at state S_i in the forward direction:

$$\delta_t(i) = P(O_1, O_2, O_3, ..., O_t, q_t = S_i|\lambda) \tag{2.46}$$

by induction:

$$\delta_t(j) = \max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right] b_j(O_t) \tag{2.47}$$

It is initialized using the initial state distribution:

$$\delta_1(i) = \pi_i b_i(O_t) \tag{2.48}$$

To keep track of the instantaneous state which has the maximum probability to transit from previous state to present state, another parameter Ψ_t is used:

$$\Psi_t(j) = \begin{cases} 0, & t = 1 \\ \arg\max\limits_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right], & \forall \quad t > 1 \end{cases} \tag{2.49}$$

Then the optimal state sequence is back-track starting from the end state $\hat{q}_T$ holding maximum probability δ_T:

$$\hat{q}_T = \arg\max_{1 \le i \le N} \left[\delta_T(i)\right] \tag{2.50}$$

$$q_t^* = \Psi_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, ..., 1 \tag{2.51}$$

It is worth noting that the HMM model does not explicitly incorporate any duration information. The self transition probability a_{ii}, which represents statistical duration information, is

used to determine the duration of being in the same state. For any non-zero a_{ii}, the result is a geometrical distribution, which is equivalent to the duration being a one-time step. Also, the transition probability remains the same no matter at what instance it is calculated. This increases the chances of misclassification of an instantaneous state in the sequence where it is expected to have the same state. Therefore, HMM poorly suits for the segmentation of signals where duration information is important.

In order to incorporate duration information, duration probability 'p' is introduced as an additional parameter in the standard HMM. The duration dependent hidden semi-Markov model (HSMM) is defined as:

$$\hat{\Lambda} = \{A, B, \pi, p\} \tag{2.52}$$

where $p = \{p_i(d)\}$ is the probability density function of being in state i along duration d. Based on the inclusion of new parameter, the Viterbi algorithm is modified as under:

$$\delta_t(j) = \max_d \max_{i \neq j} \left[\delta_{t-d}(i) a_{ij}\right] p_j(d) \prod_{s=0}^{d-1} b_j(O_{t-s}) \tag{2.53}$$

For HSS, d is considered as the entire duration of a heart cycle [5]. Both duration of being in current state and previous state that maximize (2.53) are vital information and they are tracked at each instant of time t:

$$\Delta_t(j) = \arg\max_d \max_{i \neq j} \left[\delta_{t-d}(i) a_{ij}\right] p_j(d) \prod_{s=0}^{d-1} b_j(O_{t-s}) \tag{2.54}$$

$$\Psi_t(j) = \arg\max_{i \neq j} \left[\delta_{t-\Delta_t(j)}(i) a_{ij}\right], \qquad 1 \leq i \leq N \tag{2.55}$$

The most likely state sequence is obtained by using the extended back-tracking algorithm [5]. This algorithm extrapolates the state likelihood δ_{t-d} beyond the beginning or end of a PCG sequence enabling to model the state sequence at the start or end of the PCG signal. This is achieved by equating the parameters beyond the beginning or end of the sequence to

the respective parameter value at the start or the end point of the sequence. These changes are introduced to (2.53) and (2.54) by equating $\delta_{t-d} = \delta_1,\ \forall\ t < d$ and $\delta_{t-d} = \delta_T,\ \forall\ t > T$. Similar changes are made in the observation sequence too. The extended algorithm is given as:

$$T^* = \arg\max_t \left[\delta_{T:T+d_{max}-1}(j)\right] \qquad 1 \leq j \leq N$$
$$q^*_{T^*} = \arg\max_{1\leq j\leq N} \left[\delta_{T^*}(j)\right]$$
$$t = T^*$$

Perform the following back-tracking algorithm to get the desired optimal state sequence:

$$\left.\begin{aligned} d^* &= \Delta_t(q^*_t), \\ q^*_{t-d^*:t-1} &= q^*_t, \\ q^*_{t-d^*-1} &= \Psi_t(q^*_t), \\ t &= t - d^*, \end{aligned}\right\} \quad \forall\ t > 1 \tag{2.56}$$

2.4.1 HSMM for segmentation of PCG

In [5,16], the HSMM model for HSS is built over two assumptions. First, the state sequence will always occur in a fixed order, i.e. $\underbrace{\mathrm{S1} \rightarrow \mathrm{systole(Sys)} \rightarrow \mathrm{S2} \rightarrow \mathrm{diastole(Dia)}}_{\text{one heart cycle}} \rightarrow \mathrm{next\ heart\ cycle}$. Secondly, the duration of each component (state) is consistent for an individual subject. Discussions of how the duration parameters are estimated is presented in chapter 5. These assumptions simplify the initialization of model parameters and improves the segmentation accuracy. The transition matrix $\{a_{ij}\}$ and the initial state probability $\{\pi_j\}$ need not be trained. It also avoid misclassification of an observation in the expected location of any standard heart cycle.

Both these assumptions are ideal for almost all the heart sound signals, both normal and pathological conditions. Certain abnormal conditions, such as skipped heartbeat (also known as heart palpitations) or heart rhythm problem (arrhythmias), may fail the model due

to the presence of irregularities in the heart rate. The algorithm will wrongly predict the expected state in a pre-programmed location even though it does not occur in the signal. Such misclassification may easily occur in a PCG signal exhibiting large heart rate variation (HRV). A healthy subject may also experience HRV during inspiration or expiration (also termed as respiratory sinus arrhythmia), deep breathing during medical examination, or because of physical or mental stress. Sometimes the difference between the minimum and the maximum heart cycle duration (HCD) values is much larger than the average intervals. In such condition, taking a single Gaussian probability density function (PDF) as a state duration model may not fit the desired spectrum. The model is constrained by the mean μ value indicating the center of the PDF and the variance σ^2 describing the dispersion around it. If a single-mode distribution has to cover the full extent of such data distribution, the soft gradient of the model will not match the distribution of the actual data points. The estimated likelihood using the model will not be accurate.

A multi-modal duration distribution based HSMM can be introduced to address this problem. The model will be a cascade of weighted Gaussian distribution of all possible state duration of a subject. That way each mode acts as a local search space with a sharper gradient converging the likelihood towards the expected duration. The final state duration is estimated as the instance of the distribution which bears the maximum likelihood value.

2.5 Performance matrices

The performance of heart sound segmentation is evaluated based on the ability of the algorithm to locate the S1 and S2 sounds. The reference positions of the S1 and S2 sounds used for validation are the locations of R-peak and the end of T-wave in the corresponding ECG, respectively. A tolerance of $\pm$ 40 ms is permitted for the onset and offset of events to match the ground truth. An S1 sound is labelled as truly positive (TP) if the start of the S1 sound segment is found within 40 ms from its reference point. The same goes for S2 sounds. The heart sounds detected at the wrong location are labelled as false positive (FP). And

those that are not detected at the intended reference positions are labelled as false negative (FN). Then the performance is measured using the F_1 score:

$$F_1 = 2 \cdot \frac{P_+ \cdot \mathrm{Se}}{P_+ + \mathrm{Se}} \tag{2.57}$$

In the equation, Se is sensitivity (or recall) and (P_+) is positive predictivity (or precision). They are calculated as:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\,\% \tag{2.58}$$

$$P_+ = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\,\% \tag{2.59}$$

The F_1 score gives a harmonic mean of sensitivity and precision which is a better measure of incorrect classification than the accuracy measure [94]. In recent literature, this matric is commonly used to evaluate the performance of the HSS algorithm [5, 6, 63, 69, 95].

On the other hand, in the PhysioNet/CinC Challenge 2016 for classification of normal/abnormal heart sound recordings, modified sensitivity (Se) and specificity (Sp) are defined [4, 18] for evaluating the performance.

$$\mathrm{Se} = \frac{\mathrm{wa}_1 \cdot \mathrm{Aa}_1}{\mathrm{Aa}_1 + \mathrm{Aq}_1 + \mathrm{An}_1} + \frac{\mathrm{wa}_2 \cdot (\mathrm{Aa}_2 + \mathrm{Aq}_2)}{\mathrm{Aa}_2 + \mathrm{Aq}_2 + \mathrm{An}_2} \tag{2.60}$$

$$\mathrm{Sp} = \frac{\mathrm{wn}_1 \cdot \mathrm{Nn}_1}{\mathrm{Na}_1 + \mathrm{Nq}_1 + \mathrm{Nn}_1} + \frac{\mathrm{wn}_2 \cdot (\mathrm{Nq}_2 + \mathrm{Nn}_2)}{\mathrm{Na}_2 + \mathrm{Nq}_2 + \mathrm{Nn}_2} \tag{2.61}$$

where wa_1 and wa_2 are the percentages of good signal quality and poor signal quality PCG signals available in the abnormal category, respectively. On the other hand, wn_1 and wn_2 denote the percentages of good signal quality and poor signal quality PCG signals available in the normal category. The other notations are denoted as in Table 2.2. The overall challenge score is calculated as the average value of Se and SP.

$$\mathrm{M\_Acc} = \frac{(\mathrm{Se} + \mathrm{Sp})}{2} \tag{2.62}$$

These same matrices will be used in this thesis work to evaluate the performance of the

Table 2.2: Notations for determining the modifier performance matrices [4].

Reference label	Weights	Classifier output		
		Abnormal	Unsure	Normal
Abnormal (good)	wa_1	Aa_1	Aq_1	An_1
Abnormal (poor)	wa_2	Aa_2	Aq_2	An_2
Normal (good)	wn_1	Na_1	Nq_1	An_1
Normal (poor)	wn_2	Na_2	Nq_2	Nn_2

N.B. An_1 indicates a good quality abnormal signal wrongly classified as a normal signal.

segmentation and classification of heart sound signals.

2.6 Motivation of this thesis work

Heart sound is the direct consequence of hemodynamic and rhythemic cardiac events in association with the myocardium and heart valves. It carries acoustic information that relates with the structural and functional integrity of the heart. As a result, PCG siganl can be used as diagnostic feed for computational evaluation and identifying cardiac anomalies. The study on the mechanism of heart sound production has ascertained that pathological anomalies are reflected in PCG signal as extra gallop sounds or murmurs. The morphology and spectral properties of these anomalies are distinguishable and can be used as diagnostic features for identifying their pathological relations. For example, murmurs associated with VHD are mostly high-frequency rasping sound that may last for the entire systole or diastole intervals. The temporal locations of these murmurs are unique and implicate the valves that produce them. The motivation of this thesis work is to explore various signal processing tools to establish the diagnostic features of heart sound signals. The features can be used for the segmentation of heart sound or for detection of pathological murmurs.

In the last decade, denoising of PCG signals is carried out by simply using BPF or wavelet transform. The filters work well with noise whose frequency content is outside the targeted frequency range. The methods may not be able to suppress the noise frequency components in the systole and diastole intervals that overlapped with the normal heart sound signals. On

the other hand, the nonlinear TVF is widely used to restore piecewise constant signal. It is a smoothing filter that estimates the expected signal that retains the essential information of the signal in the form of first-order derivatives. The process removes any highly varying low lying noise producing a smooth discontinuous region. The overlapping group sparsity (OGS)-based TVF adopts the structural sparsity of the first order derivatives resulting in a slower transition of the signal waveform. This resolves the problem of the stair-case artifact. PCG signals fit with the profile of the signals that are applicable for implementing the TVF. Motivated from the prospects discussed above, a hybrid denoising method has been proposed that combine both the LTI filter and the TVF. First, the BPF is applied to remove any high-frequency noise. Then, the residual low-intensity noise will be smooth out by TVF. The degree of smoothing required for each signal is determined using a data-dependent approach by measuring the complexity of the signal.

The FHS are the prominent sounds that are easily detectable in PCG signal. Accurate detection of these two signals helps estimate the duration of the cardiac cycle and locate the systole and diastole intervals. Therefore the envelope feature is used extensively for the analysis of PCG signal. Typically, Shannon entropy and Shannon energy are calculated to uniform intensify the envelope of FHS. But these methods have their limitations. As an alternate method, the logistic function is investigated. The idea is to map the signal amplitudes of the FHS to the upper asymptote of the S-curve leaving the noise signals towards the asymptotic tail. The motivation of this method is to adjust the growth rate and the center of the S-curve according to the intensity level of the FHS and the noise. With proper calibration, the transformed waveform will uniformly enhance the FHS envelop peaks which segregate from the systole and the diastole intervals with a large amplitude margin.

Heart sound segmentation is the foremost step in any CAD system. Among the state-of-the-art algorithms, HSMM and bidirectional recurrent neural networks (BiRNN) that analyse the duration dependency of the signal are well known. As reported, both the algorithms yield equivalent performance score for segmentation of the S1, systole, S2 and diastole. The BiRNN based HSS algorithm is relatively versatile and easily extended to the detection of

other cardiac events. But, the training process requires a large dataset. HSMM based HSS algorithm works better with a small dataset. Due to the use of a single-mode duration model, the existing HSMM based algorithm is not able to accurately segment the PCG with variable heart rate. A large variation in heart rate demands for the variable duration model for each state. The problem may be resolved by introducing a multi-mode duration model covering all possible state durations in a PCG signal. Each mode represents local duration space where the actual state duration is expected. The multi-mode model provides a sharper gradient of the likelihood that helps estimate the final state duration more accurately using the maximum likelihood criterion.

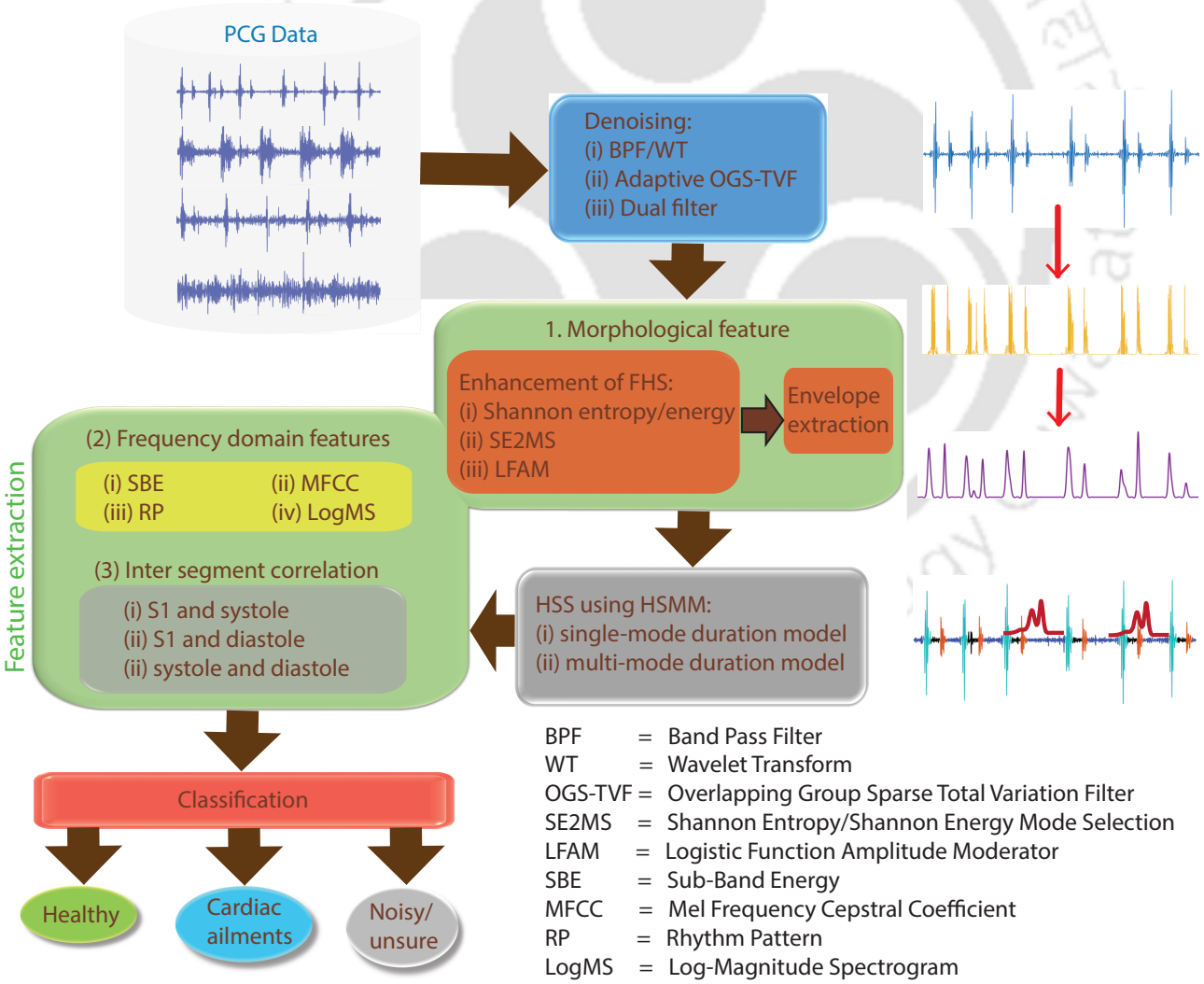

Figure 2.1: Schematic diagram showing the plan of the proposed investigations in this thesis work.

To capture the tone and timbre of sounds, the distribution of character frequencies in the spectrogram are analysed at different subbands. Among the frequency domain features, features such as MFCC, SBE, and log-magnitude spectrogram are popularly used for heart sound analysis. But, there is no definite subband feature suitable for the classification of murmurs yet. It is still an open problem, and there is a scope for exploring new subband features. For classifying PCG signals into broad categories of normal, murmur, and noisy, the dominant frequency ranges of each category can be explored for feature extraction. Murmurs due to VHD usually occur in either systole or diastole intervals. While noise generated from the recording device or the ambient are mostly continuous and last all through the signal. This relation may also be explored as features. With this motivation, a new subband feature is proposed. In addition, the correlation coefficients of pairs, S1 vs systole, S1 vs diastole and systole vs diastole are also proposed. Since the signal waveform with the same spectral property may not be identically distributed in the time domain, the correlation coefficient values are calculated from the proposed SBE feature of each segment.

Based on the above discussions, the proposed investigations in this thesis are planned as follows. The schematic diagram of the work plan is shown in Fig. 2.1.

- To design a hybrid denoising method that combines both the LTI filter and the TVF: In TVF, the required degree of smoothing is determined using a data-dependent approach based on the complexity measure of the signal.

- Develop an alternate method to enhance the detection of FHS using the logistic function: It involves adjusting the growth rate and the center of the S-curve according to the signal intensity levels of the FHS and the noise. The calibration will automatically adjust the amplitudes of the FHS to the upper asymptote of the S-cure and the noise signals towards to asymptotic tail of the S-curve.

- Proposes a better heart sound segmentation algorithm: A multi-mode duration model is introduced to the HSMM to improve the versatility of the classifier against the variable state duration. The model is made signal-dependent by estimating the model for

individual signals. The number of modes in the model depends on the variation of state duration in the PCG signal.

- To classify PCG recording into broad categories of normal, murmur, and noisy: It investigates the dominant frequency ranges of each category and derives a new subband feature. In addition, the correlation coefficients of pairs, S1 vs systole, S1 vs diastole and systole vs diastole, are also explored.

3

Denoising heart sounds using dual filtering approach

Contents

Preprocessing is the foremost step in any signal analysis process. The heart sound signal or phonocardiogram (PCG) is a signal that is prone to noise interference. The real-time PCG signals are affected by artifacts or noise picked up from the data recording device or the surrounding as discussed in Chapter 2 Section 2.1. Therefore, the preprocessing module is essential before proceeding with heart sound segmentation (HSS) and heart sound classification (HSC). This Chapter evaluates the adaptive total variation filter (TVF) and its modified extensions for denoising a PCG signal. In Section 2.2, a brief highlight of TVF for PCG denoising is introduced. The method has a clear potential to improve the denoising process if appropriate smoothing parameters are provided. The generic method using the overlapping group sparse regularization avoids unnecessary clipping of the output waveform.

This chapter investigates the potential aspects of TVF and the possible extensions that will help improve the performance of PCG denoising. The study indicates how the noise level affects the choice of smoothing parameter (λ) value and thus the denoising process. As reported, a high noise level requires a larger λ value. Therefore, as part of the extension work, the λ value is derived as a function of signal complexity. This makes the method signal specific enabling the smoothing process to adjust according to the nature of the noise in the signal. The parameter value is calculated in every iteration step until the optimization function converges to its minimizer. But TVF as a denoising tool for PCG signal has one major limitation. When the noise intensities are higher than the heart sound signal, most of the large derivative values (governing the regularization function of TVF) are from the noise signal. Because of this, the smoothing process may go in favour of noise signals. In this case, the algorithm will not be able to remove the noise effectively. To solve this issue, a hybrid dual filtering process is proposed. This process combines the band-pass filter (BPF) or wavelet transform (WT) with the TVF. The BPF/WT filter will remove the high-frequency out-of-band noise, while the TVF will smooth out the weaker residual noise.

3.1 Total variation filtering for PCG denoising

The PCG signal denoising has been motivated largely by the total variation regularization problem [1, 56, 83, 88, 96]. Among the contemporary methods, the wavelet transform (WT) based denoising is the common one [50, 78, 80, 81]. The performance is comparatively better. But the wavelet approach depends solely on the base wavelet function that governs the overall performance. A fixed wavelet may not necessarily fit all the PCG signal waveforms which come from a wide range of individuals having different age, gender, and health condition [1]. On the other hand, the TVF is emerging as a good denoising method for PCG signals. It is solved as an optimization problem by imposing constrain on the sparsity of the first-order derivatives, or total variation (TV) of a signal. It effectively smooths out the background noise by retaining the dominant TV values. The standard TVF often results in staircase artifacts that are suitable only for piecewise-constant signals. The repercussion of this method is a major limitation when executing it on any other signal that may be locally approximated by higher-order polynomials and exhibit group sparsity of large derivative values. PCG signal is one such example that has large derivative values clustered at the locations of S1 and S2 sounds and sparsely distributed over the remaining segments. The signal has the nature of approximately piecewise-smooth and does not discontinue abruptly but extends over some intervals. To accommodate more variety of signals, a generic method known as overlapping group sparse (OGS) based TVF has been developed [1, 88]. The method calculates the cumulative effects of the neighbouring fluctuation in the signal and consequently smoothes the signal. A detailed discussion of OGS based TVF is included in Chapter 2.

The smoothing parameter λ used in both the variance of TVF (standard and OGS based) are empirically set. Such an approach is not appreciable when the noise in the signal may have different SNR levels. In an effort to develop an adaptive algorithm, Deng et al. [1] exploited a penalty function derived from hyperparameter assigned as Gamma prior and using Bayesian inference. The value is calculated iteratively via the maximum a-posteriori (MAP) estimation method. The method is comparatively stable. But introducing Gamma-prior

in the penalty function produces additional shape and scale parameters. Though these parameters have little effect on denoising process, they still need predetermining before execution. In addition, the method uses noise variance as the main component to derive the smoothing parameter value and as a stop condition parameter to terminate the iteration process. But estimating the noise variance using the MAD rule is not accurate and may lead to erroneous results.

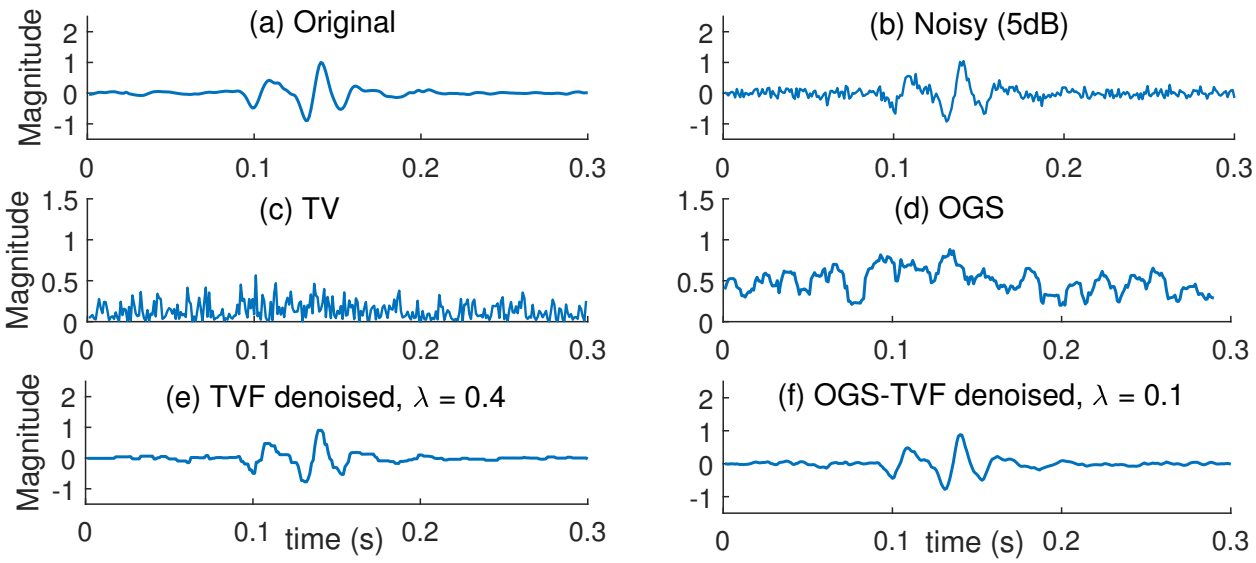

Figure 3.1: Example of (a) noise-free segment of S1 sound signal, (b) noise corrupted signal, (c) first-order derivatives or total variation, (d) group sparse total variation, and denoised signal resulting from (e) TVF and (f) OGS-TVF.

An example of the denoising process using the standard TVF and OGS-TVF on a segment of S1 sound is shown in Fig. 3.1. The noisy signal in the example is corrupted by AWGN at 5 dB SNR. To highlight the distribution of regularization function involved in both the methods, Fig. 3.1(c) and (d) show the absolute value of the first-order difference, $|\mathbf{Dx}|$ and overlapping group total variation, $\|\mathbf{Dx}_{n,K}\|_2$ for each of the solutions. Total variation promotes the sparsity of $|\mathbf{Dx}|$ which on solving produces sharp changes at the location of large derivative values based on the penalty weight. The process preserves the discontinuities in the signal. When implemented on a more generic piecewise smooth signal, the sharp transition nature of TVF introduces a staircase artifact on the signal. This is observed in the denoised signal illustrated

in Fig. 3.1(e). The waveform shows small flat regions rather than a progressive smooth surface. The OGS-TVF, shown in Fig. 3.1(d), gradually adjust the large derivative values to neighboring values by calculating cumulative sparse derivatives. Doing so, it tracks the progressive changes of the signal rather than abrupt changes. The denoising process, Fig. 3.1(f), substantially reduces the staircase behavior. In this example, the group size is set at $K = 10$;

To evaluate the quality of the denoised signal, the signal-to-filter error ratio (SFER) is measured [1, 56, 97]. Given a clean (noise-free) PCG signal $y(n)$ of length N and the filter output $f(n)$, the SFER is calculated as

$$\text{SFER} = 10\log_{10}\frac{\sum_{n=1}^{N} y^2(n)}{\sum_{n=1}^{N}\left(y(n) - f(n)\right)^2} \tag{3.1}$$

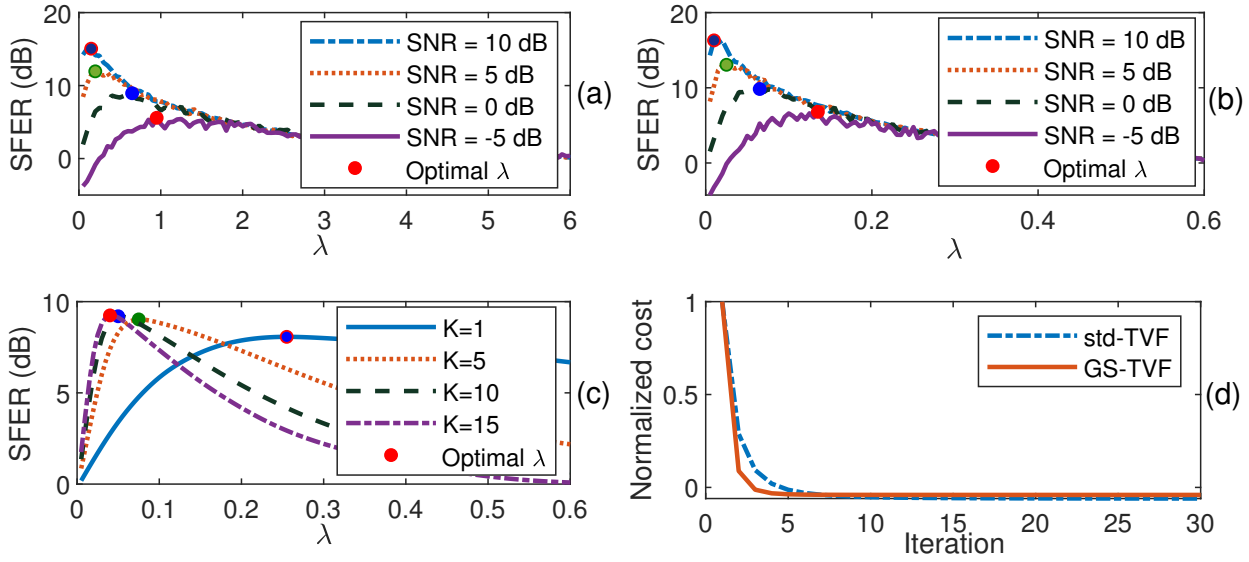

Figure 3.2: SFER measuring the quality of denoising for signal corrupted with AWGN noise and taking different λ values for (a) TVF and (b) OGS-TVF with group size $K = 10$, (c) OGS-TVF with different group size when noise SNR is 0dB. (d) Show the convergence rate of both the method.

Fig. 3.2 shows the SFER values for different smoothing parameter (λ) values. The optimal λ value for both the denoising problems are affected by noise levels, see Fig. 3.2 (a) and (b). Therefore there is a need to derive an adaptive λ value that will effectively denoise a given noisy signal. In the case of OGS-TVF, the group size also affects the denoising

performance. Small group size results in a sharper transition while large size (say $K = 20$) results in smoother transition through longer intervals. This is analogous to a low pass filter. Introducing a larger group size reduces staircase artifacts. Comparatively, the OGS-TVF needs a smaller λ value and converges faster than the standard TVF.

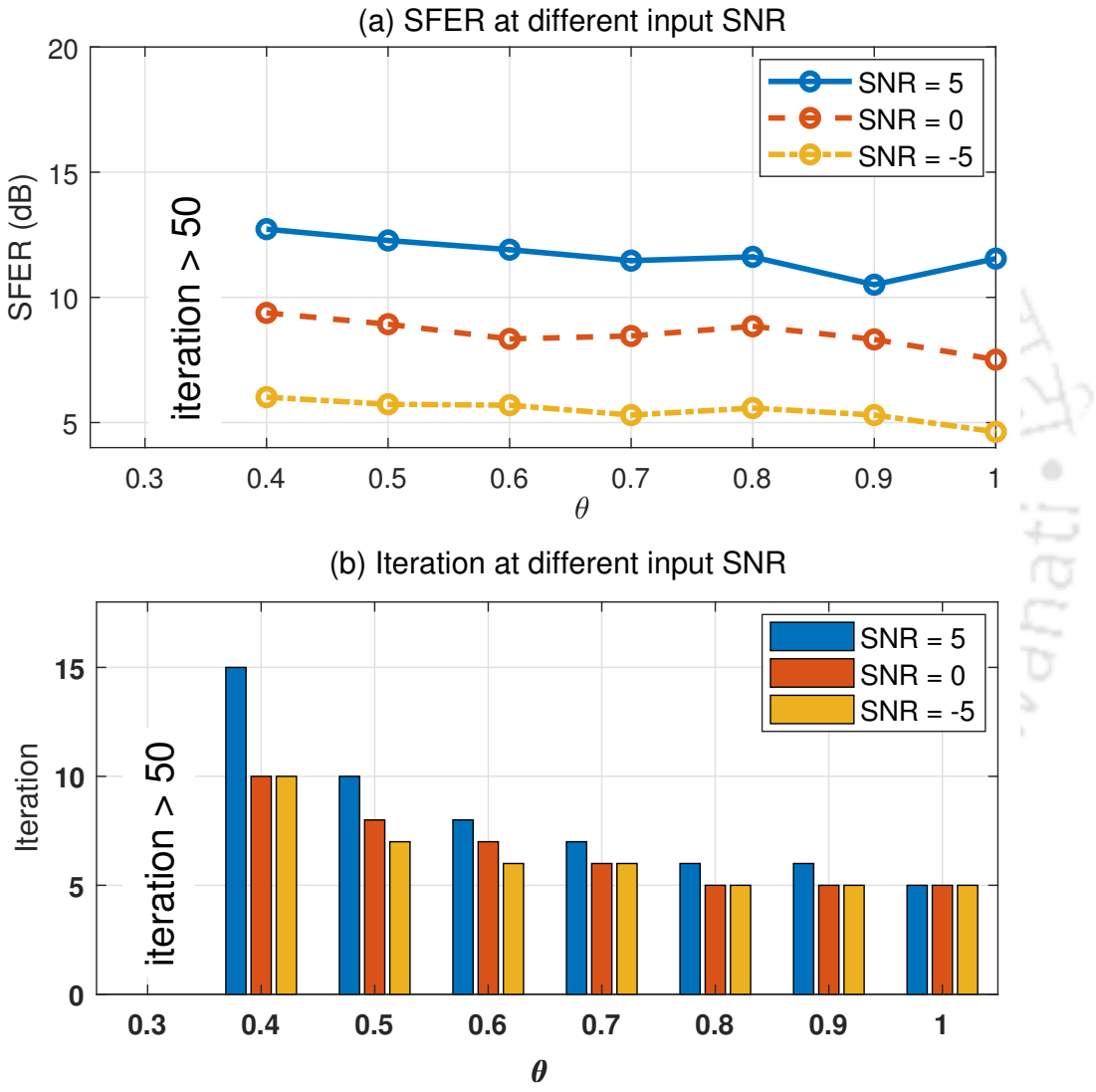

Figure 3.3: The SFER metric and the number of iterations at input SNR = 5, 0, -5, the θ value lies between 0.4 and 1 at steps of 0.1. In panels (a) SFER metric at different input SNR levles and (b) iterations required at different input SNR levels.

Gamma prior based smoothing parameter $\rho(\mathbf{v}) = \frac{\varrho^2(\Theta N+\dot{\alpha})}{\varphi(\mathbf{D}\mathbf{v}+\dot{\beta})}$ in adaptive OGS-TVF [1] is capable of denoising PCG signal adaptively. But this method leads to additional parameters such as θ, $\dot{\alpha}$ and $\dot{\beta}$. On solving Eq. (2.26), $\dot{\alpha}$ and $\dot{\beta}$ have little effect on denoising performance

which may even be neglected. The constant θ scales the derived λ value affecting the degree of smoothing. A small θ gradually smooths the signal at smaller steps of increasing penalty function values improving the denoising accuracy. The process may need more iteration steps for convergence. Large θ value increases the difference ($|\lambda_{i+1} - \lambda_i|$) which improves the convergence rate. Due to the larger stepping of λ_i, the filtering process may not give an optimal result. The SFER and the number of iteration steps of the MAP-based OGS-TVF when taking different θ values and for different noise levels are shown in Fig. 3.3 (a) and (b). For $\theta < 0.4$, the method becomes too expensive.

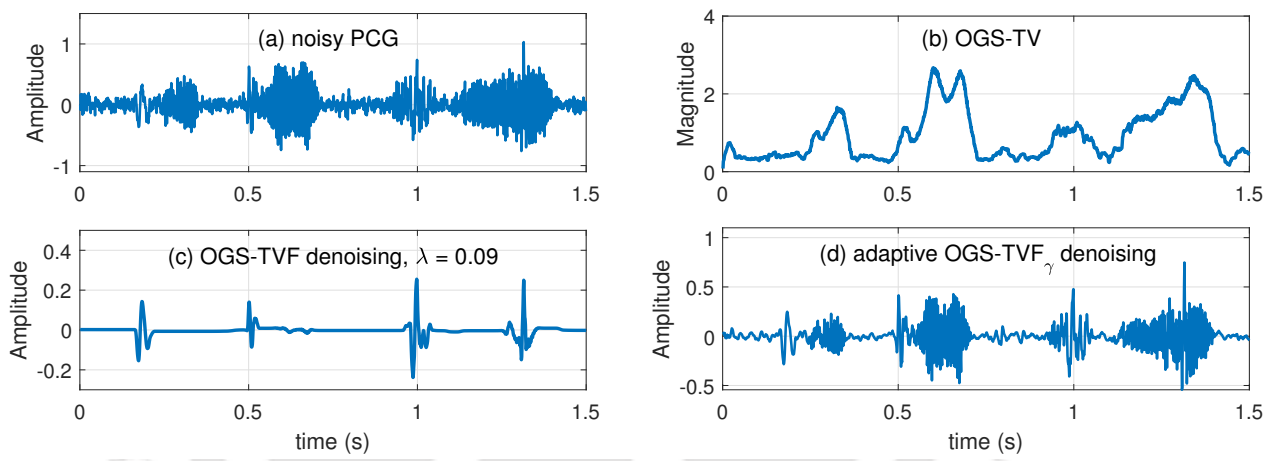

Figure 3.4: Example of OGS-TVF denoising of a noisy signal that is locally corrupted with noise and with noise intensity higher than the actual heart sound signal.

Since the algorithm is built on the assumption that noise is identically distributed across the signal, it is feasible only for denoising only incessant noise. When noise is locally introduced in the signal, like respiratory noise or voices, the adaptive penalty function $\rho(\mathbf{v}) \propto \frac{1}{\varphi(\mathbf{Dv})}$ fails to recognize them as noise. Fig. 3.4 illustrates an example of PCG signal with a voice and the denoised signal using OGS-TVF with $\lambda = 0.09$ and adaptive OGS-TVF [1]. If a suitable smoothing parameter is set, the OGS-TVF can remove the local noise as shown in Fig. 3.4 (c). The result of adaptive OGS-TVF is shown in Fig. 3.4 (d). It can eliminate Gaussian noise, but the local noise (voices) remains unaffected.

Fig. 3.5 illustrates the λ_i value derived using MAP estimation at each iteration steps. The λ_i value gradually increases with each iteration step until it reaches its optimal value of

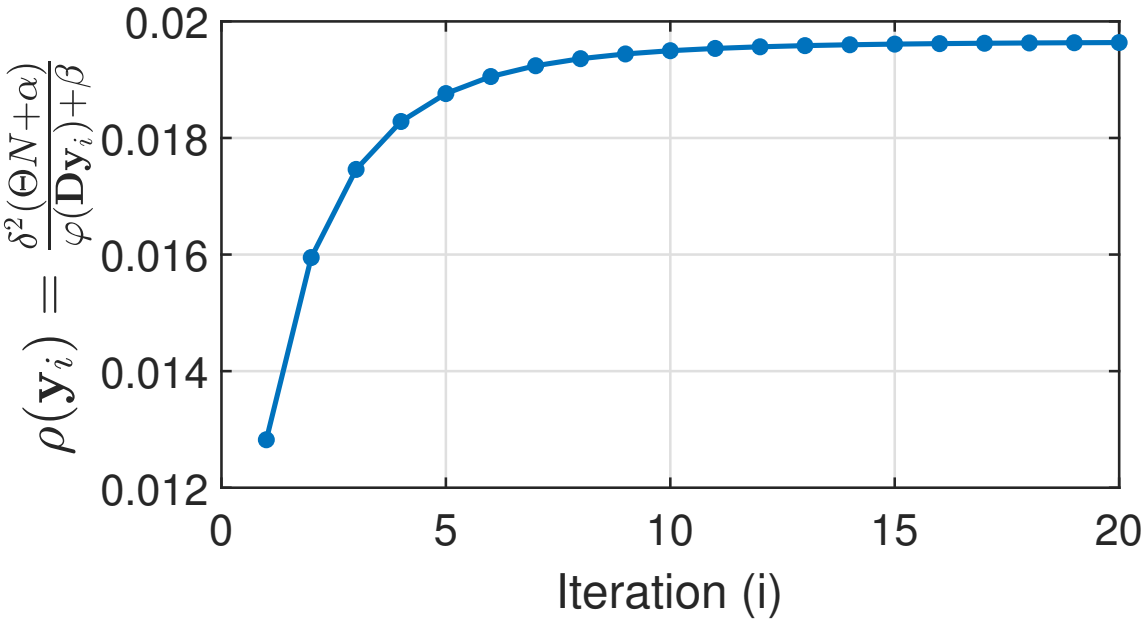

Figure 3.5: λ values of adaptive OGS-TV denoising algorithm [1] at each iteration steps.

≈ 0.02, observed after 10 iterations. This is because the estimation of $\rho(\mathbf{v}) \propto \frac{1}{\varphi(\mathbf{D}\mathbf{v})}$ produces ascending λ value until optimal value is achieved. Sometimes this can be a problem. Taking an increasing λ_i value with each step may overprocess the already smooth signal waveform. Without a softer smoothing process, the final output may not converge accurately.

3.1.1 Proposed adaptive penalty function

The proposed penalty function attempts to incorporate the signal information in terms of the complexity measure of the signal. In TVF, the λ value controls the degree of smoothing performed on a given signal. Therefore, determining this value based on the signal complexity measure can be an effective adaptive method. Among other measures reported in the literature, sample entropy (SampEn) is widely used to assess the complexity of the time-series signal. It is calculated from a signal sequence $x(n) = x(1), x(2), ..., x(N)$ by creating a smaller template vector, $\mathbf{x}_{n,K} = [x(n), ..., x(n+K-1)]$, of size K. This produces $(N-K+1)$ template vectors in total. Then the probability of template pairs that are matching is calculated

as

$$\Phi^K = \frac{\underset{l\neq n}{\text{count}}\left(d[\mathbf{x}_{l,K},\mathbf{x}_{n,K}]<r\right)}{N-K+1}, \qquad l,n = 1:N-K+1 \tag{3.2}$$

where r is a noise filter parameter. In general $r = \text{tolerance} \times \text{std}$. The sample entropy is calculated as

$$\text{SampEn}(K,r,N) = \ln \frac{\Phi^K}{\Phi^{K+1}} \tag{3.3}$$

The advantage of using this entropy measure is that it reduces the biasing due to self-matching, and it is relatively consistent at different sequence length [98, 99]. Therefore, to reduce the computational cost, $\text{SampEn}(2, 0.4 \times \text{std}, \text{HCD})$ is calculated from one heart cycle duration (HCD) of a PCG signal. The smoothing parameter λ is defined using the sample entropy measure as:

$$\lambda = \theta \times \text{SampEn}(2, 0.4, \text{HCD}) \tag{3.4}$$

where $\theta \in (0,1)$ is a constant. It is set at 1 for the standard TVF and 0.1 for OGS-TVF. Algorithm 2 shows that the updation taking place at each iteration.

Algorithm 2 Total variation denoising algotirthm.

Initialize
$\mathbf{y}_1 = \mathbf{x}$
$i = 1$
repeat
$\lambda_i = \theta \times \text{SampEn}(\mathbf{y}_i, 2, 0.4, \text{HCD})$
$\mathbf{F} = \frac{2}{\lambda_i}\Lambda_i^{-1} + \mathbf{D}\mathbf{D}^\mathbf{T}$ ($\mathbf{F}$ is tridiagonal)
$\mathbf{y}_{i+1} = \mathbf{x} - \mathbf{D}^\mathbf{T}(\mathbf{F}\backslash\mathbf{D}\mathbf{x})$ (fast solver)
$i = i + 1$
until convergence or satisfy a condition

3.1.2 Stop condition

Allowing the TVF algorithm to terminate at the convergence of the cost function may cause over smoothing of a signal. Conditioning the algorithm by equating the variance of the residual

signal $\mathbf{x} - \mathbf{y}$ to the noise variance as in [1] has a limitation when the noise signal is unknown and locally introduced into the signal. Because when calculating the variance, the local noise level is averaged along the signal length. The algorithm [1] assumes the noise variance to be known and estimates its value using the MAD rule. It does not check the actual noise variance and may cause erroneous results. Therefore, a new event using sample entropy is suggested. The stop condition is

$$\mathrm{SampEn}(\mathbf{y}_i, 2, 0.2, \mathrm{HCD}) > \frac{\mu(\mathrm{SampEn}(\mathrm{clean\_PCG}))}{\hat{f}_n} \tag{3.5}$$

where $\hat{f}_n = \frac{\max(|\mathbf{y}_i|)}{\max(|\mathbf{x}|)}$ is the factor with which the denoised signal amplitude has changed from the original signal. This condition is efficient since SampEn is less affected by sequence length.

3.2 Proposed dual filtering: LTI band-pass filter with OGS-TVF

The non-linear smoothing using total variation has some shortcomings. It does not remove the noise based on the frequency compositions. It suppressed the relatively high varying components by trading with signal amplitudes. When the noise level is too high, it causes unnecessary smoothing of the signal waveform and may lose details. To compromise between the signal quality and the noise removal, we propose a combination of TVF and an LTI filter. The LTI filter is effective against out-of-band noise. But when the signal and noise frequency bands are very close (or overlapping), the denoising process becomes challenging. The hybrid of TVF with the LTI based denoising may be a good combination to suppress the residual low-amplitude noise left out after the LTI filter. This study uses two popular LTI filters, the BPF, and the wavelet-decomposition-based filter. For BPF, the frequency range is set between 20 Hz and 150 Hz to accommodate only the FHS [18]. In the wavelet-decomposition-based filter, considering that the PCG data sampled at 1 kHz, the noisy signal is decomposed into five levels. Then from the detail coefficients, whose frequency range fall within the

bandwidth of FHS i.e. d_3 (62.5-125 Hz), d_4 (31.25-62.5 Hz), and d_5 (15.6-31.25 Hz), are considered and the PCG signal is reconstructed. Thus, we propose three different type of filter structures to denoise FHS signals.

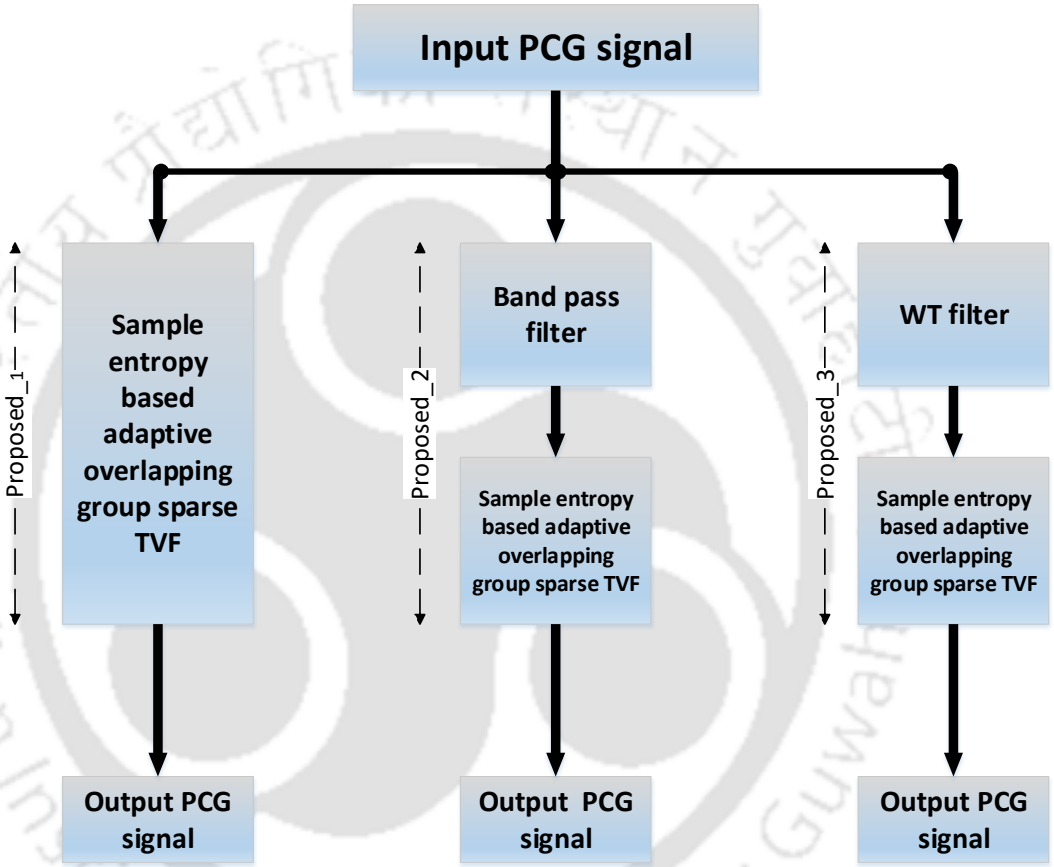

Figure 3.6: Block diagram of proposed filtering schemes.

- Sample entropy based adaptive overlapping group sparse TVF ($\mathrm{Proposed}_1$).
- Band pass filter with sample entropy based adaptive overlapping group sparse TVF ($\mathrm{Proposed}_2$).
- WT filter with sample entropy based adaptive overlapping group sparse TVF ($\mathrm{Proposed}_3$).

In Fig. 3.6, a block diagram of proposed filtering schemes is shown. In the '$\mathrm{Proposed}_1$' scheme the complexity of the time series data of PCG is calculated using sample entropy

measure which in turn decide the smoothing parameter λ of the filter. The pseudocode is shown in Algorithm 2. In the '$\mathrm{Proposed}_2$' scheme, the signal after the BPF filter is passed through the sample entropy-based adaptive OGS-TVF for better denoising. Similarly, in the '$\mathrm{Proposed}_3$' scheme, the signal is first subjected to WT based filtering. These two filtering approaches, '$\mathrm{Proposed}_2$' and '$\mathrm{Proposed}_3$' are dual filtering and they are expected to yield better denoising of the FHS signals. The performance of the denoising scheme will greatly affect the FHS envelope detection and segmentation of the heart sound.

The performance of the FHS denoising is evaluated by measuring SFER (Eq. (3.1)), and root mean square error (RMSE). The RMSE is defined as

$$\mathrm{RMSE} = \frac{\|\mathbf{y} - \mathbf{f}\|_2}{\sqrt{L}} \tag{3.6}$$

These measures are meaningful when evaluated against noise-free PCG signals. Based on the results, the parameters are regularized as standard.

3.3 PCG dataset used for evaluation

The PCG data used for evaluation is available in PhysioNet/Computing in Cardiology Challenge 2016 [18]. These data sets are described in Chapter 2 Section 2.1 of this thesis. The augmented dataset [5, 72] has 792 PCG recordings of 135 subjects, sampled at 1 kHz. Of these, 386 PCG recordings are normal, and the remaining 406 signals are abnormal. Some of these signals may be corrupted by noise signals such as stethoscope motion, intestinal and breathing sound, and the hospital environment noise, such as beeping machine sound, talking, crying, ventilation noise and many more. There is no dedicated PCG data that is not corrupted by noise. Therefore, from the dataset, fifty PCG signals that have relatively low noise levels are shorted out. Among them, some heart sound signals may have very weak S2 sounds. In this relatively noise-free dataset, we try to accommodate a wide range of normal signals with different heart rates.

3.4 Results and discussions

Evaluating the denoised signal of abnormal PCG signals may become challenging. Because there is no normal PCG recording of the patient with which the results can be compared. Based on the knowledge of signal characteristics, assuming S1 and S2 sounds are the only audible heart sound, the inter and intra sound amplitude variations of denoised signals are examined. To do that, the box plot of each sound component is analyzed and their correlations are evaluated. For adaptive OGS-TVF, the optimal regulation parameter λ value depends on the input SNR level and the group size K. In Fig. 3.2, the effect of input SNR and group size on the λ value has been shown. As observed in the figure, increasing the group size Kimproves the performance of the OGS-TVF [1] which in turn increases the computational cost. In our study, the group size is set at $K = 10$. To test the OGS-TVF incorporating the MAP estimation method by Deng et al. [1], the parameters are set accordingly as θ = 0.8, α = 1 and β = 50.

In wavelet-decomposition-based filtering, the PCG signals are decomposed using the discrete wavelet transform (DWT) with various wavelet families. The popular wavelet for heart sound analysis are Daubechies 'dB5', 'dB10', biorthogonal 'bior3.9', 'bior5.5', Coiflet families. From the detail coefficients d_3 , d_4 , and d_5, the signal is reconstructed. To determine the suitable wavelet, the envelope values from the reconstructed signal of each wavelet family are summed for S1, S2, and silent intervals. The wavelet family that yielded the highest ratio of the sum of the amplitudes for the S1 and S2 sounds compared to the sum over other intervals is selected for further use [5].

Fig. 3.7 and Fig. 3.8 illustrate the resulting filtered signal waveform using the conventional and the proposed methods at two different SNR levels, 10 dB and -5 dB respectively. The conventional BPF and wavelet decomposition-based filter is effective against the out-of-band noise signals irrespective of the noise intensity levels. When dealing with low-frequency noise that overlaps with the spectrum of heart sound signals, neither of the methods may have adequate selectivity to reject the noise. It may suppress the noise to some extent when the

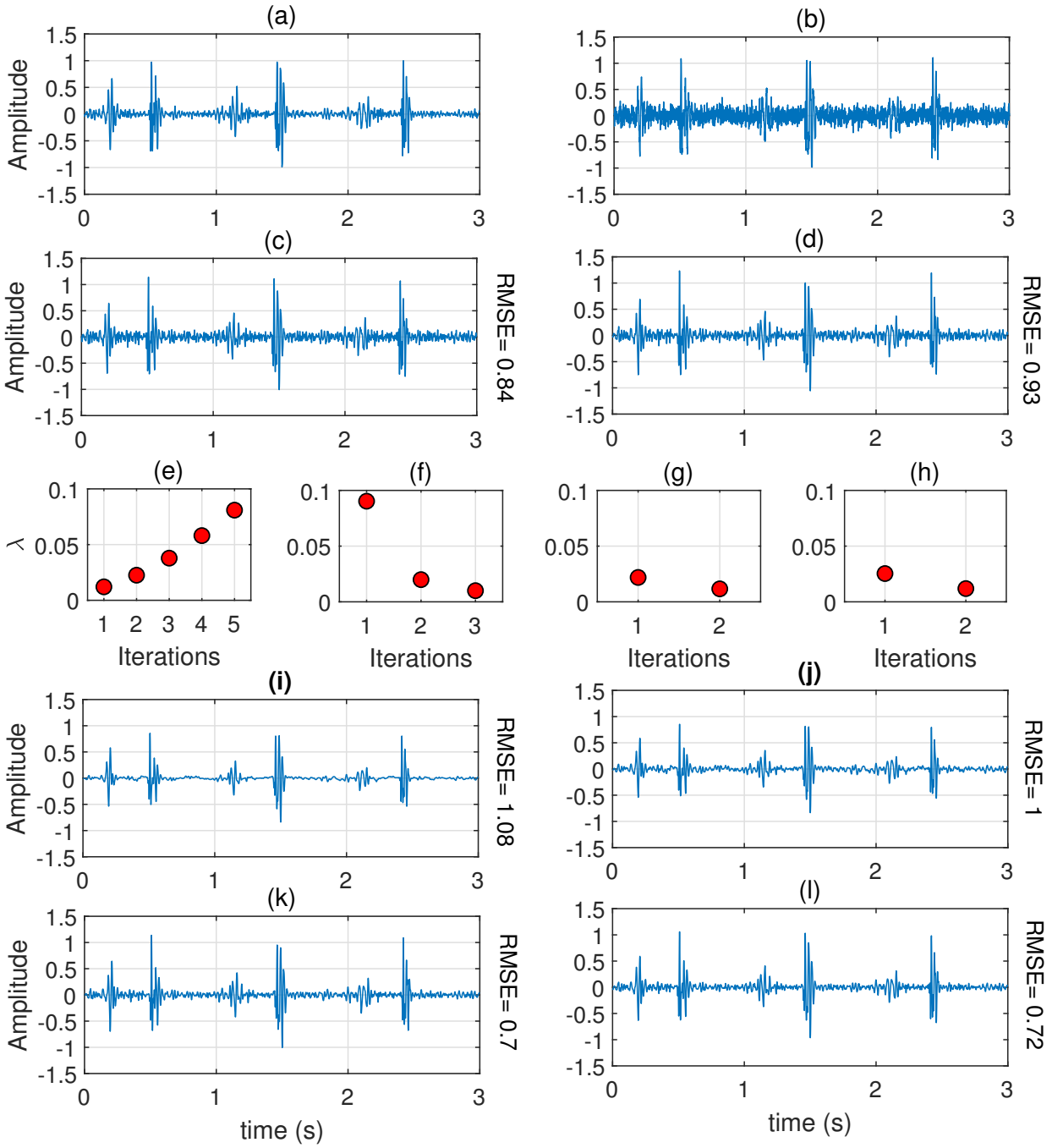

Figure 3.7: Illustrate (a) PCG signal, (b) affected by AWGN noise of 10 dB SNR. The resulting filtered signals from (c) BPF, (d) wavelet transform using'db10' (i) adaptive OGS-TVF [1], (j) proposed OGS-TVF using sample entropy and the dual filtering that combines (k) wavelet transform and (l) BPF with OGS-TVF are also shown. The λ values generated at each iteration corresponding to OGS-TVF are shown in (e), (f), (g), and (h), respectively.

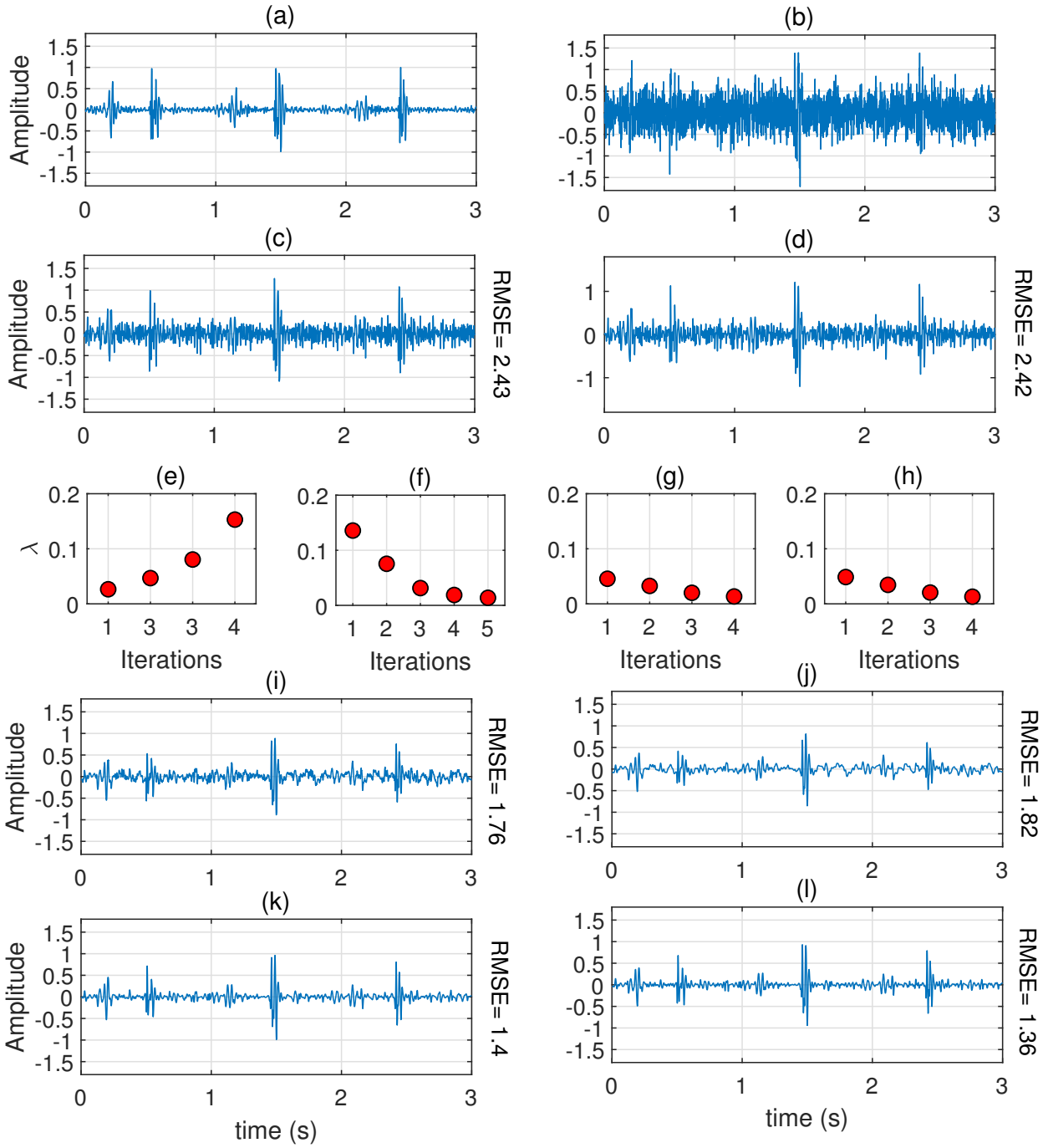

Figure 3.8: Illustrate (a) PCG signal, (b) affected by AWGN noise of -5 dB SNR. The resulting filtered signals from (c) BPF, (d) wavelet transform using'db10' (i) adaptive OGS-TVF [1], (j) proposed OGS-TVF using sample entropy and the dual filtering that combines (k) wavelet transform and (l) BPF with OGS-TVF are also shown. The λ values generated at each iteration corresponding to OGS-TVF are shown in (e), (f), (g), and (h), respectively.

SNR level is high as shown in Fig. 3.7 (c) and (d). But when the SNR level decreases, the noise interferences are not subdued adequately resulting in noisy output, as shown in Fig. 3.8 (c) and (d).

The denosing results of PCG signal using conventional group-sparse total variation filter which was suggested by Deng et al. [1] are shown in Fig. 3.7 (i) and Fig. 3.8 (i) at two SNR levels 10 dB and -5 dB respectively. The λ values generated by MAP estimation at different iteration steps are shown in Fig. 3.7 (e) and Fig. 3.8 (e) respectively. The results of denoising using this method produces better performance compared to BPF and WT filter methods. But estimating the noise variance using the median absolute deviation (MAD), oversmooth the signal and yields a large root-mean-square-error (RMSE) value 1.03 (Fig. 3.7 (i)) and 1.76 (Fig. 3.8 (i)). On the other hand, using sample entropy as regularization parameter and as the stop condition in the proposed method, $\mathrm{Proposed}_1$, improves the denoising accuracy as shown in 3.7 (j) and Fig. 3.8 (j).

Combining LTI filter with sample entropy-based adaptive OGS-TVF filter compensates for the limitations of each filter. The BPF or wavelet-based filter helps remove high-frequency out-of-band noise. This will remove most of the local noise such as voices, murmurs, or ambient noise (Fig. 3.4). The remaining noise is suppressed using the proposed adaptive OGS-TVF algorithm taking smaller λ values and lesser iteration steps (Fig. 3.7 (g), (h) and Fig. 3.8 (g), (h)). The process also avoids waveform distortion of the denoised signal. The results of denoising yield a smaller root-mean-square-error (RMSE), 0.7 and 0.72 for 10 dB SNR. The plots shown in Fig. 3.7 (k) and Fig. 3.8 (k) are results of dual filtering with wavelet-transform based filter and the proposed adative OGS-TVF. Similarly, Fig. 3.7 (l) and Fig. 3.8 (l) are the results of dual filtering using BPF and the proposed adaptive OGS-TVF. Both the hybrid methods give comparable results. For easy implemention and computatinal cost, the BPF with OGS-TVF is preferrable.

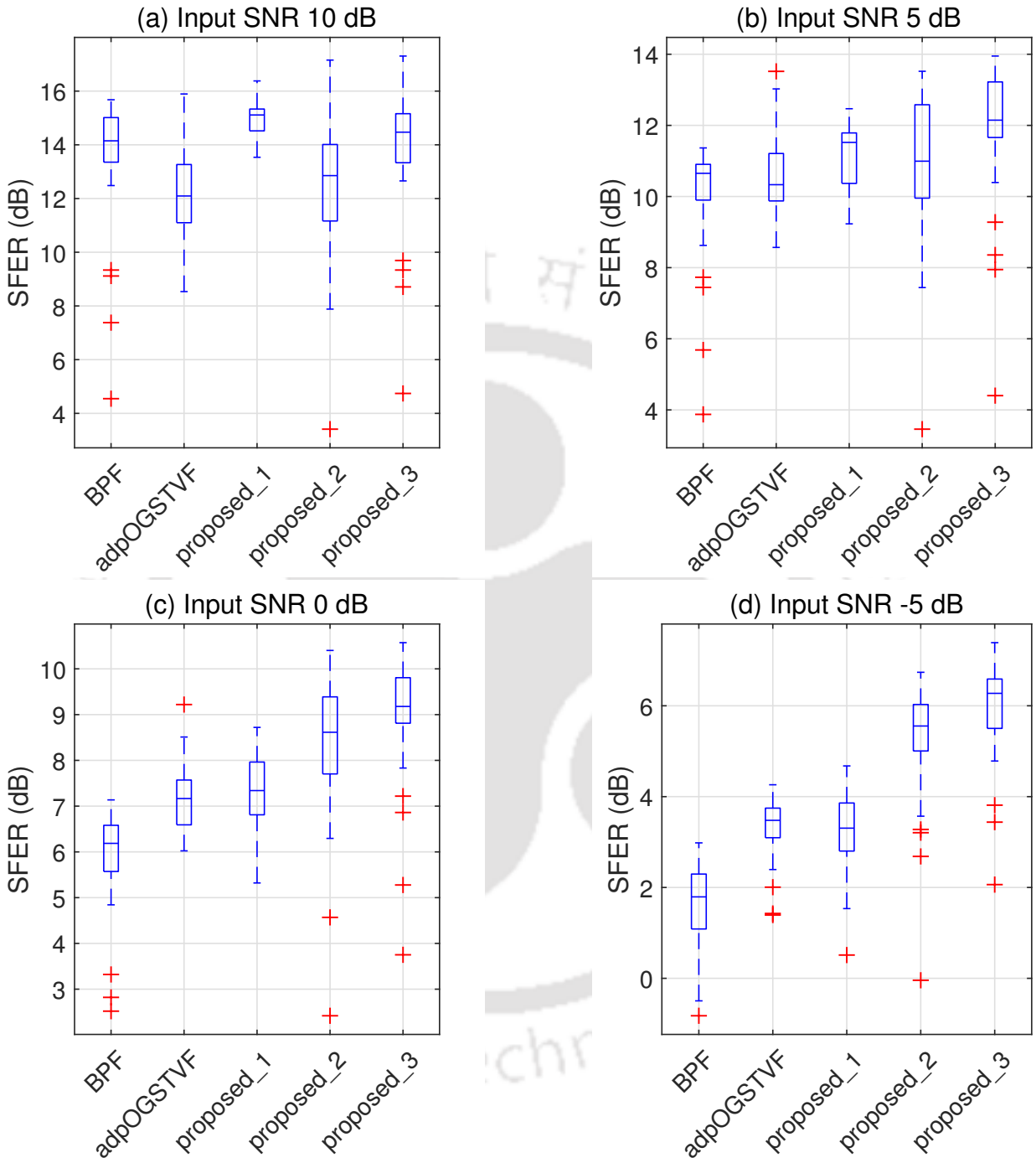

Figure 3.9: The signal-to-filter-error ratio (SFER) for different input noise.

Fig. 3.9 shows the signal to filter error ratio of the denoising methods, BPF, adaptive OGS-TVF, '$\mathrm{Proposed}_1$', '$\mathrm{Proposed}_2$' and '$\mathrm{Proposed}_3$' at different SNR levels. At higher SNR input signal (10 dB), the dual filtering approach ('$\mathrm{Proposed}_2$' and '$\mathrm{Proposed}_3$') gives SFER values around 13 and 14 respectively. For lower SNR value (-5 dB), the dual filtering apprach

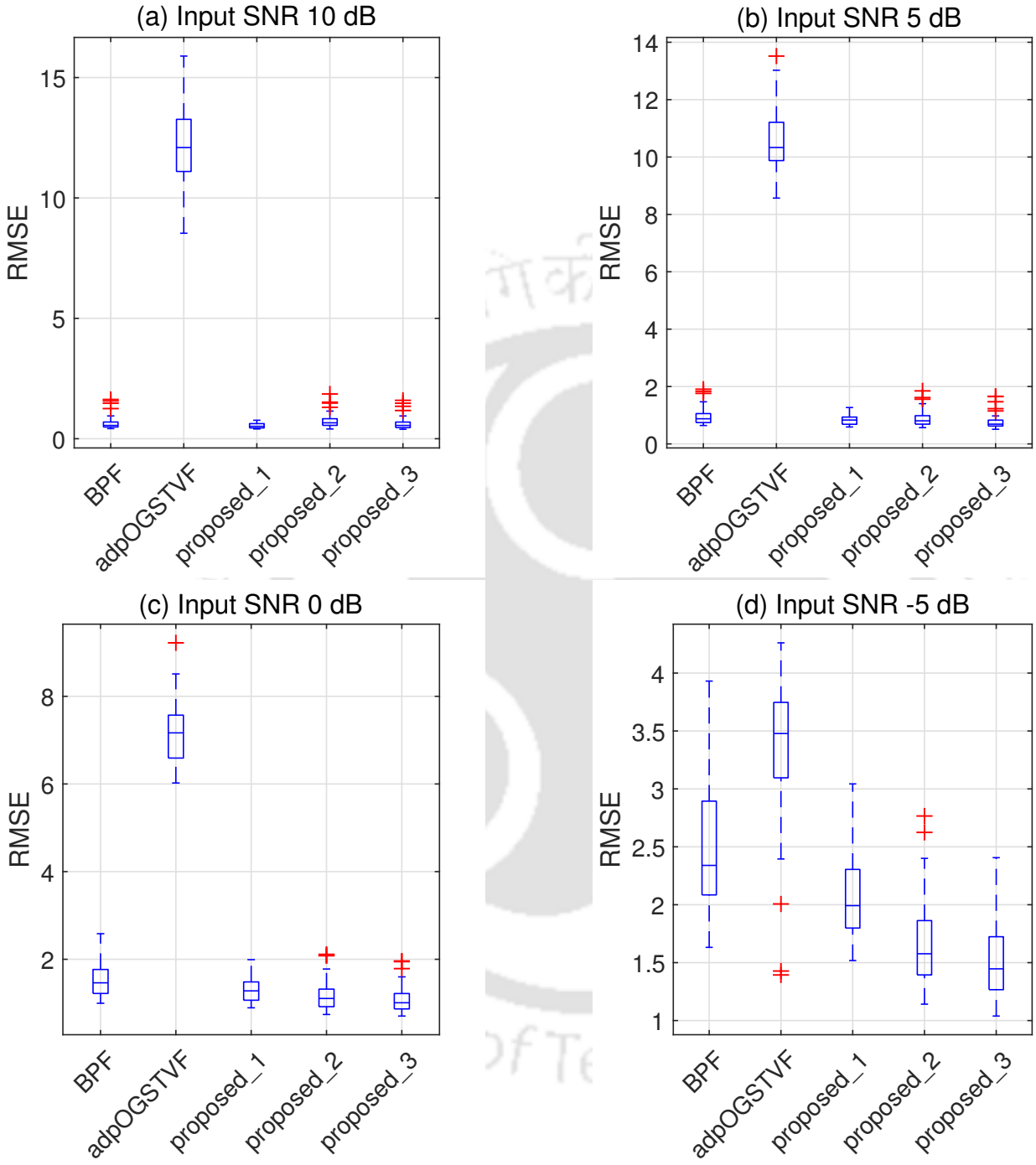

Figure 3.10: The root-mean-square error (RMSE) for different input noise.

gives better SFER value between 5 to 6. In Fig. 3.9 (a), (b), (c) and (d), the SFER values are shown for different input SNR levels, 10 dB, 5 dB, 0 dB and -5 dB for conventional BPF, adaptive OGS-TVF and proposed methods respectively. The proposed dual filtering approaches give better performance in terms of SFER metric.

The performance of the filters in terms of RMSE are measured for comparison and shown in Fig. 3.10. For the large input SNR, the proposed methods, '$\mathrm{Proposed}_1$', '$\mathrm{Proposed}_2$' and '$\mathrm{Proposed}_3$' yield better performance than the conventional methods. In Fig. 3.10 (a), (b), (c) and (d), the RMSE values are shown for different input SNR levels, 10 dB, 5 dB, 0 dB and -5 dB respectively for conventional BPF, adaptive OGS-TVF and proposed methods. Comparison of the RSME metric also reveals that the dual filtering approaches give better performance for heart sound signals.

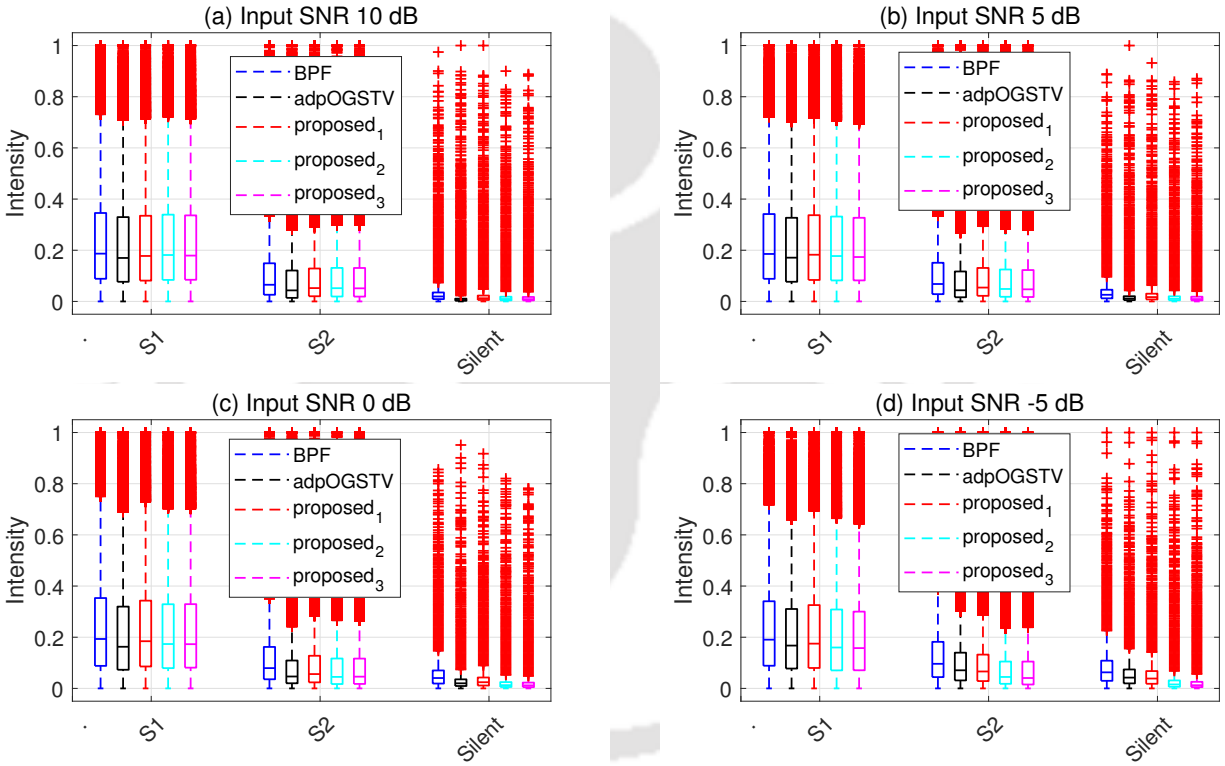

Figure 3.11: Intensity distribution of FHS signals.

For analysing the FHS (S1 and S2 sounds), the denoising process should preserve their signal characteristics. Therefore, the proper signal envelope detection and accurate FHS segmentation process depend also on the quality of the denoising process employed. So, the inter and intra sound amplitude variations of denoised signals are needed to be examined. To do that intensity of each sound component is analyzed in the box plot and their dominance in the denoised signals are evaluated in Fig. 3.11. Analysis of signal intensity distribution of different heart sound components at different SNR levels, 10 dB, 5 dB, 0 dB and -5 dB are

shown in Fig. 3.11 (a), (b), (c) and (d) respectively. In all the cases it is seen that the intensity of FHS is well preserved by our proposed denoising schemes. Even the silence periods are well maintained with accurate signal intensity. This has established the performance of the proposed methods.

3.5 Summary

This Chapter presents an adaptive OGS-TVF algorithm for denoising of heart sound signals. The existing adaptive OGS-TVF algorithm utilizes the MAP estimation of the original signal based on hierarchical Bayesian inference. The algorithm does not need predefined smoothing parameter values and can denoise the signal adaptively. But it works on the assumption that the noise is independent and identically distributed where the common variance is presumed to be known. The statistical MAD rule is used to estimate the noise variance in the algorithm. The estimated noise variance value is also used in the pre-stop condition to decide the termination of the iteration steps. Therefore, the performance of heart sound denoising using this algorithm depends on how accurate the MAD rule estimates the noise variance. When evaluated for the denoising performance using SFER and RMSE measures, the algorithm shows a high error rate.

In the proposed adaptive OGS-TVF algorithm, signal complexity measures known as the sample entropy is estimated from the individual signal to determine the smoothing parameter. By calculating the sample entropy value, the noise level in the signal is assessed. A high sample entropy value indicates a high noise level in the signal. Therefore deriving smoothing parameter value and the pre-stop condition using the sample entropy measure in the proposed algorithm is more effective.

Based on the proposed adaptive OGS-TVF algorithm, two new hybrid schemes for denoising are proposed. The methods combine the LTI filter (BPF and WT based) with the OGS-TVF to efficiently denoise PCG signals. In the dual/hybrid filtering methods, the inclusion of the BPF or the wavelet-transform-based filter improves the band rejection of out-of-band

noise so that the TVF will further smooth out the residual weaker noise in the signal. A stop-condition can be devised to skip the TVF denoising and prevent over smoothing the signal by analyzing the sample entropy measure. On comparing the hybrid filtering methods with other existing methods, the hybrid methods show considerable improvements over others.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

4

Enhancement of heart sound envelope

Contents

In this Chapter, we will discuss a new method to extract the envelope of the fundamental heart sounds using logistic function. The basic sigmoid function, also known as logistic function, is the key component of this logistic function amplitude moderation (LFAM) method. The proposed LFAM involves finding critical amplitudes, also known as lower and upper cut-off amplitudes. These critical amplitudes are dependent on the nature and the degree of noise contained in the signal. Their values are regressively obtained from the signal itself by histogram analysis of intensity distribution. We also proposed a Shannon entropy and a Shannon energy based amplitude moderation of the heart sound to extract its envelope. This method selects either Shannon entropy or Shannon energy looking at the signal.

4.1 Heart sound intensities and logistic function

The signal intensities of the fundamental heart sounds (FHS), S1 and S2, have been viewed as key features for the analysis of heart sound signals [5, 100]. This concept originates from the perceptual evaluation of heart sound carried out in the clinical environment. During auscultation, the S1 and S2 sounds are heard as the dominant intensity sounds separated by a silent systolic or diastolic interval in every cardiac cycle. By analyzing the intensity information along with tone and timbre, clinical experts can identify these sounds. For this reason, we proposed a logistic function based amplitude moderation, where the logistic function is parameterized by introducing the scaling parameter α and shifting parameter β. These correspond to the upper cut-off intensity level above which the signal is emphasized and the lower cut-off intensity level below which the signal will be attenuated as silent sound intervals.

Mathematically, the intensity information is defined as the envelope of a signal. Some of the conventional envelope extraction methods are absolute value, energy (by squaring), the Hilbert envelope, the homomorphic envelope, and the Teager-kaiser energy (TKE) envelope [5, 6, 34, 38]. The heart sound envelope ideally has two peaks in every heart cycle duration (HCD), indicating the intensity of S1 and S2 sounds. For the uniformity and easy detection

of FHS, it is preferred that these envelope peaks maintain more-or-less identical intensity levels. But, realizing it in the real-time application is challenging. Depending on the site of auscultation, physique or health condition of the subject, the S1 sound may appear louder than the S2 sound, or vice versa. The popular methods to mediate such inconsistency use Shannon entropy or Shannon energy-based envelogram techniques [34, 36, 37]. In these methods, the descending logarithmic weights of the amplitudes emphasize the medium intensity signals as valuable information. The Shannon entropy accentuates the lower medium intensities and makes the envelope peaks substantially uniform. Since this method emphasizes the weaker intensity signal, the envelope may become noisy. On the other hand, the Shannon energy envelope is less affected by noise. It emphasizes higher medium intensity signals and suppresses the lower intensity signals. This method reduces the effect of weaker intensity noise and maintains minimal variation between envelope peaks.

Both methods, Shannon entropy and Shannon energy, have their own advantages and disadvantages. If either of the methods is adopted properly by examining the feasibility of the method with the nature of a given signal, it can improve the detection of FHS in noisy or pathological recordings. To implement this, the system needs to identify whether a PCG signal is clean, noisy, or pathological. There is one major drawback in calculating the Shannon entropy and Shannon energy that is often overlooked. In both the methods, the signal intensities nearing its absolute maximum value are extenuated. In other words, the higher signal intensities are not considered for the candidacy of heart sounds. This results in undesired errors in the derived envelope. On the other hand, these methods do not suppress the low-intensity noise signal as desired. Instead, they equally emphasize the intensity distribution of low and medium amplitude signals, prominently seen in the Shannon entropy-based envelope. The resulting envelope appears noisy. To tackle this limitation, a logistic function based amplitude moderation (LFAM) method is proposed. This is motivated by the sigmoid-curve characteristic of the logistic function. It involves categorizing the levels of signal intensities belonging to either FHS or noise signal categories. Then the LFAM is applied to uniformly enhance all the signal intensities belonging to FHS. With proper calibration of

parameters, the remaining signal intensities are suppressed implicating the silent intervals. In the proposed method, the logistic function is parameterized by introducing the scaling parameter α and the shifting parameter β. The α and β values are calibrated according to the upper cut-off intensity level above which the signal is emphasized and the lower cut-off intensity level below which the signal will be attenuated as silent sound intervals, respectively. In this chapter, we will discuss how to obtain the optimal parameter values from a given PCG.

4.2 Logistic function amplitude moderation (LFAM)

The generalized expression of the sigmoid (S) curve logistic function is:

$$\sigma(x) = \frac{L}{1 + e^{-\alpha(x - x_0)}} \tag{4.1}$$

where, L is the maximum value of the curve; x_0 is the location of the curve mid-point; α is the steepness or growth rate of the logistic curve [101]. In literature, the logistic function has been used in various applications and in different forms. There are two popular modes of expression [102–104]:

$$\sigma(x) = \frac{1}{1 + e^{\beta_1 - \alpha x}} \tag{4.2a}$$

or,

$$\sigma(x) = \frac{1}{1 + \beta_2 e^{-\alpha x}} \tag{4.2b}$$

where, α controls the growth rate (scaling), and β_1 (= αx_0) or β_2 (= e^{β_1}) provides the shifting of the curve's mid-point.

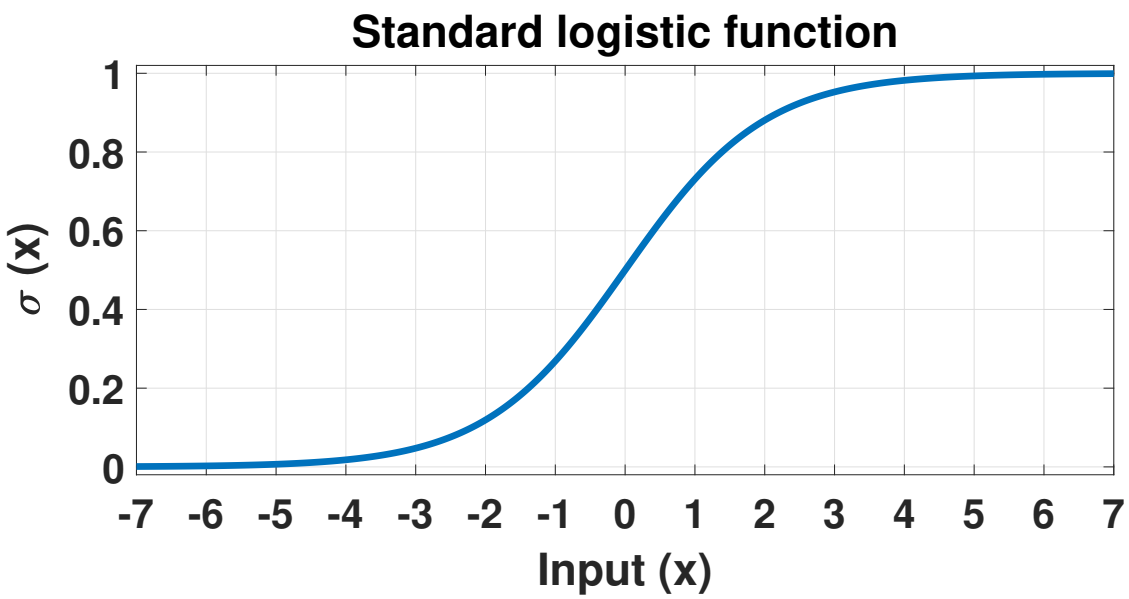

Figure 4.1: The standard logistic function generates the approximate sigmoid curve for input x ranging between -2π and 2π.

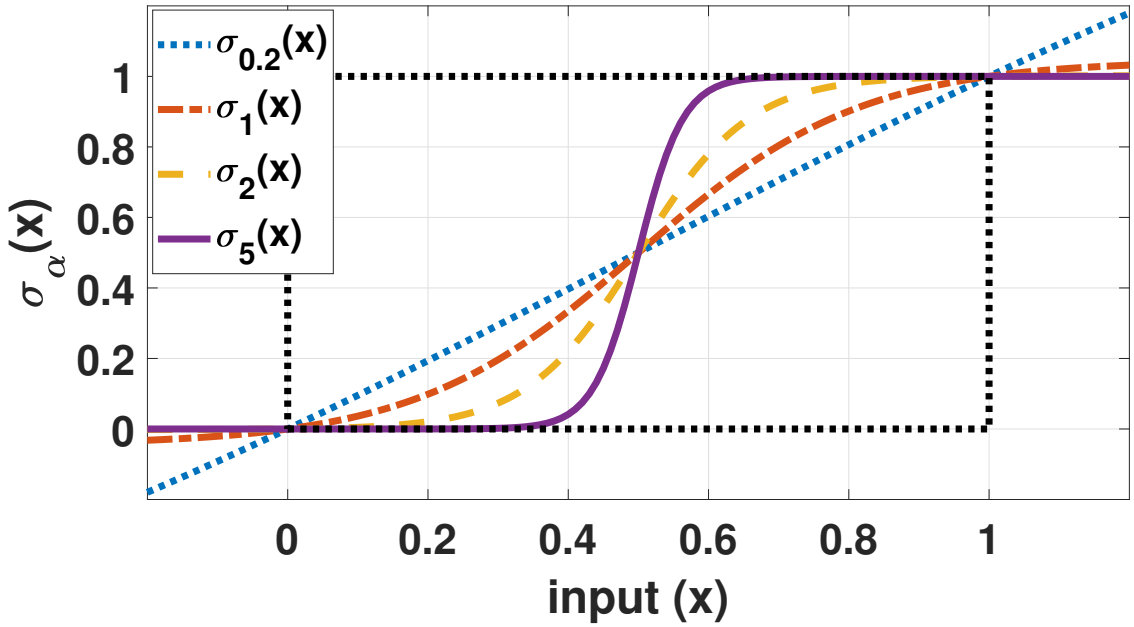

Figure 4.2: The logistic function transformation $\sigma_\alpha(x) \in [0,1] \quad \forall \quad x \in [0,1]$ at different values of α (0.2, 1, 2, 5) keeping the center of sigmoid at amplitude level 0.5.

For standard parameter values (L = 1, x_0 = 0, α = 1), the expression becomes:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4.3}$$

$$= \frac{e^{x}}{e^{x}+1}$$
$$= \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2}\right)$$

The use of the exponential function (e^{-x}) enables the standard logistic function to obtain the sigmoid curve characteristic that converges to its saturation values between 0 and 1 for x over a small range of [-2π,2π], shown in Fig. 4.1. This is ideal for various signal processing and machine learning applications. In order to obtain a conditional output ($\sigma \in$ [0,1]) of a normalized input ($|\mathbf{x}| \in$ [0,1]), irrespective of α and β (shown in Fig. 4.2), the function has to be min-max normalized.

$$\sigma(x) = \frac{\sigma(x) - \sigma(x_{\min})}{\sigma(x_{\max}) - \sigma(x_{\min})} \tag{4.4}$$

The Eq. (4.2a) and (4.2b) becomes:

$$\sigma(x) = \frac{\left[1 - e^{-2\pi\alpha x}\right]\left[1 + e^{\beta_1 - 2\pi\alpha}\right]}{\left[1 + e^{\beta_1 - 2\pi\alpha x}\right]\left[1 - e^{-2\pi\alpha}\right]} \tag{4.5a}$$

or,

$$\sigma(x) = \frac{\left[1 - e^{-2\pi\alpha x}\right]\left[1 + \beta_2\, e^{-2\pi\alpha}\right]}{\left[1 + \beta_2\, e^{-2\pi\alpha x}\right]\left[1 - e^{-2\pi\alpha}\right]} \tag{4.5b}$$

They are further approximated as:

$$\sigma(x) = \frac{\left[1 - e^{-2\pi\alpha x}\right]}{\left[1 + e^{\beta_1 - 2\pi\alpha x}\right]} \tag{4.6a}$$

or,

$$\sigma(x) = \frac{\left[1 - e^{-2\pi\alpha x}\right]}{\left[1 + \beta_2\, e^{-2\pi\alpha x}\right]} \tag{4.6b}$$

such that

$$\left.\begin{aligned} e^{-2\pi\alpha} &\leq \varepsilon \\ \Rightarrow \alpha &\geq 1 \end{aligned}\right\} \tag{4.7}$$

$$\left.\begin{aligned} e^{\beta_1 - 2\pi\alpha} &\leq \varepsilon \\ \beta_2\, e^{-2\pi\alpha} &\leq \varepsilon \end{aligned}\right\} \tag{4.8}$$

where, ε ($\approx e^{-2\pi}$ or 0.0019) is a very small value. The expressions in Eq. (4.8) can be further simplified as $\alpha \geq \frac{1}{1-x_0}$.

4.3 Projection of parameters in terms of signal amplitude

Eq. (4.6b) is used as the basic function of our proposed LFAM. It will generate a moderated output $\mathrm{M\_signal}$ generated from an input signal segment $\mathbf{x}$ of length N.

$$|\mathbf{x}| = [|x(1)|, |x(2)|, |x(3)|, ..., |x(N)|] \qquad \in [0, 1] \tag{4.9}$$

$$\mathrm{M\_signal}(n) = \sigma_{\alpha,\beta}(n) = \frac{1 - e^{-2\pi\alpha|x(n)|}}{1 + \beta\, e^{-2\pi\alpha|x(n)|}}, \qquad n = 1 : N \tag{4.10}$$

The function parameters α and β determine the degree of moderation, and their values are closely associated with the intensity levels of different signal components. Taking x_{uc} as the upper cut-off amplitude above which the signal belongs to loud FHS signals, x_{lc} as the lower cut-off amplitude below which the signal will be discarded as noise/silent intervals, and projecting them into sigmoid curve such that:

$$\sigma_{\alpha,\beta}(x_{\mathrm{uc}}) = 0.9 \tag{4.11a}$$

$$\sigma_{\alpha,\beta}(x_{\mathrm{lc}}) = 0.1 x_{\mathrm{lc}} \tag{4.11b}$$

the relation between the parameters and the critical amplitudes are derived as:

$$\frac{0.1\, e^{2\pi\alpha x_{\mathrm{uc}}} - 1}{0.9} = \frac{(1 - 0.1\, x_{\mathrm{lc}})\, e^{2\pi\alpha x_{\mathrm{lc}}} - 1}{0.1\, x_{\mathrm{lc}}} \tag{4.12a}$$

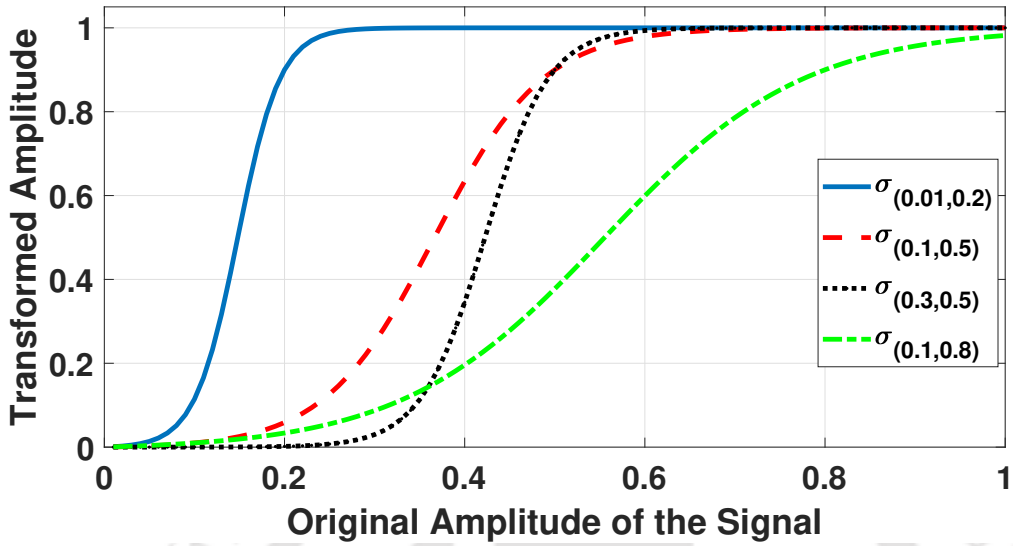

Figure 4.3: Example of LFAM based intensity distribution $\left(\sigma_{(x_{\rm lc},x_{\rm uc})}\right)$ at different lower $(x_{\rm lc})$ and upper $(x_{\rm uc})$ cut-off amplitude.

$$\beta = \frac{0.1\ e^{2\pi\alpha x_{\rm uc}} - 1}{0.9} \tag{4.12b}$$

From the roots of Eq.(4.12a), consider only the smallest real root satisfying the condition $\alpha > 1$. The example of LFAM transformation for linear intensity distribution at different critical cut-off values is shown in Fig. 4.3. This illustrates that above the specified $x_{\rm uc}$, the intensities become consistent. By correlating this value with the lowest detectable heart sound intensities and applying along the whole length of a signal, the transformed waveform can produce uniformity of heart sound envelope peak intensities. The figure also shows that the intensities below $x_{\rm lc}$ are attenuated. If this value is estimated to the possible noise intensities, then it may suppress noise in a PCG envelope.

4.4 Estimation of lower and upper cut-off amplitudes

It is very crucial to determine the values of $x_{\rm uc}$ and $x_{\rm lc}$. With admissible values, the derived S-curve will improve the segregation of FHS from noisy systole and diastole. The resulting LFAM signal will have sharper envelope peaks and smoother silent intervals. If these values are not discriminative, the proposed method is not better than the existing methods and may

even cause more distortion. For the effective categorization of the signal intensities, the critical amplitudes are regressively obtained from the signal itself. It is clear that the values of x_{uc} and x_{lc} are largely dependent on the nature of the signal. For a relatively clean PCG signal with distinct silent systole and diastole, both x_{uc} and x_{lc} will be small. In the case of a noisy signal, the parameter values solely depend on the noise level. Therefore, the intensity level of FHS and noise are first examined by using the intensity-based histogram analysis. The histogram analysis is carried out by considering the bin size as '$\frac{1}{R}$', where R is the number of intensity bins. Then the distribution functions (DF) are calculated.

The distribution function of signals below intensity level '$\frac{i}{R}$':

$$\mathrm{DF}_{\mathrm{low\_amp}}(i) = P\left[|\mathbf{x}| \leq \frac{i}{R}\right], \qquad i = 1:R \tag{4.13}$$

The distribution function of signals above intensity level '$\frac{i}{R}$':

$$\mathrm{DF}_{\mathrm{high\_amp}}(i) = P\left[|\mathbf{x}| > \frac{i}{R}\right], \qquad i = 1:R \tag{4.14}$$

Then the signal information preserved in the loud heart sound signals after suppressing the noise intensities ($\leq \frac{i}{R}$) is estimated as the relative value of the sum of amplitudes:

$$\{x'_i(l')\}_{l' \in [1, L' \leq N]} = \left\{|x(n)| : |x(n)| > \frac{i}{R}\right\}_{n \in [1,N]}, \quad i = 1:R \tag{4.15}$$

$$\mathrm{DF}_{\mathrm{SoA}}(i) = \frac{\sum_{l'=1}^{L'} x'_i(l')}{\sum_{n=1}^{N} x(n)} \tag{4.16}$$

Lastly, the degree of loudness (not in dB scale) of a given signal sample above intensity level '$\frac{i}{R}$' is defined as the square of its mean amplitude.

$$\mathrm{DF}_{\mathrm{loudness}}(i) = \left(\frac{1}{L'}\sum_{l'=1}^{L'} x'_i(l')\right)^2 \tag{4.17}$$

$$\mathrm{DF}_{\mathrm{loudness}} = \mathrm{DF}_{\mathrm{loudness}} - \mathrm{DF}_{\mathrm{loudness}}(1) \tag{4.18}$$

The Eq. (4.18) gives the relative degree of loudness keeping the baseline as the mean of total signal intensities. Using these parameters, the upper cut-off amplitude x_{uc} is re-defined

as the amplitude that maximizes the number of signal samples in the category of loud sound:

$$x_{\mathrm{uc}} = \max_i \left(\mathrm{DF}_{\mathrm{high\_amp}}(i) \cdot \mathrm{DF}_{\mathrm{loudness}}(i)\right) \times \frac{1}{R} \tag{4.19}$$

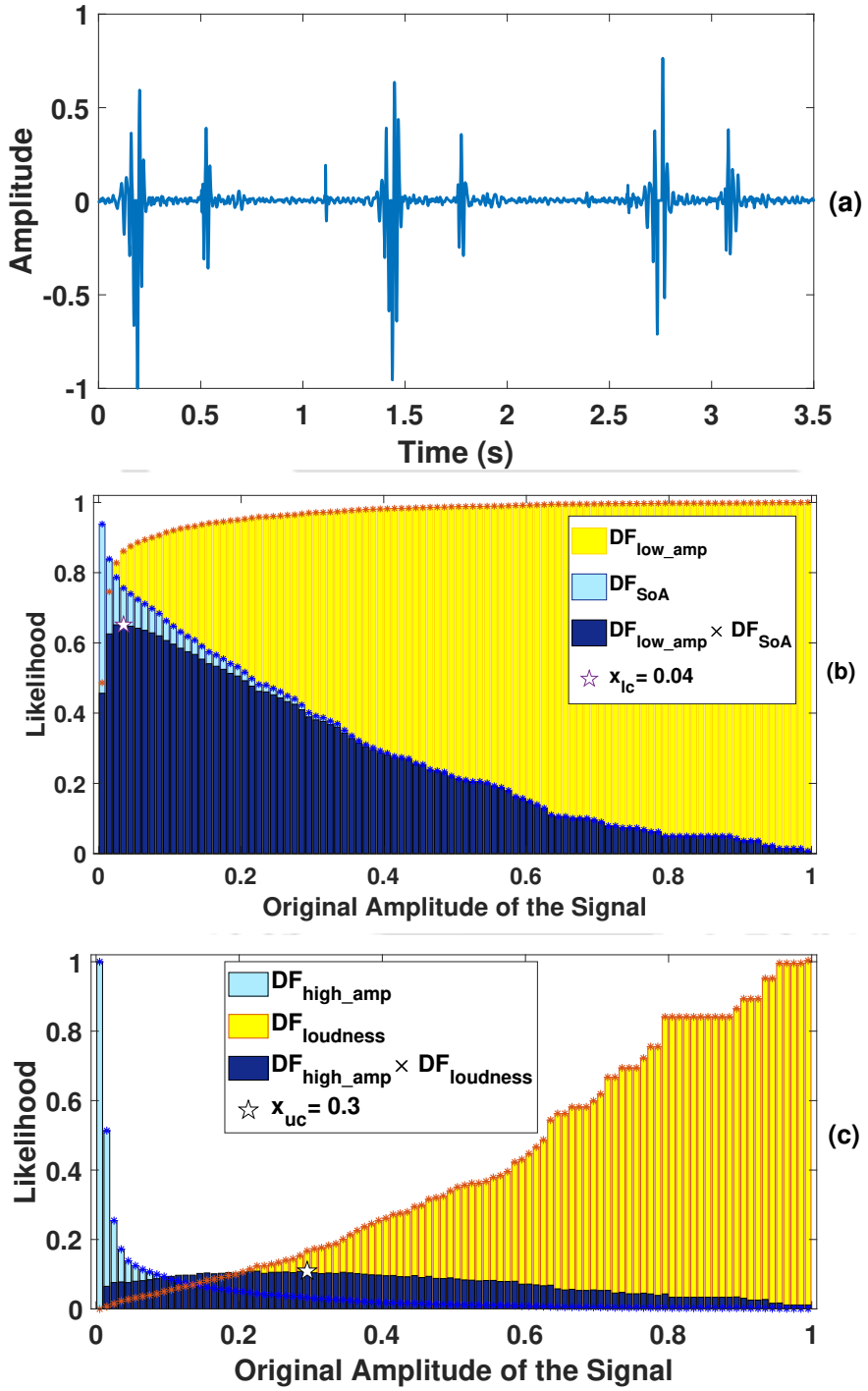

Figure 4.4: Example of (a) a normal PCG, and the corresponding steps to calculate (b) lower cut-off amplitude x_{lc}, and (c) upper cut-off amplitude x_{uc}.

The lower cut-off amplitude x_{lc} is also defined as the amplitude that maximizes the number of signal samples in silent sound level, and in due time retain most of the signal information in the remaining louder signal intensities.

$$x_{\mathrm{lc}} = \max_{i} \left(\mathrm{DF}_{\mathrm{low\_amp}}(i) \cdot \mathrm{DF}_{\mathrm{SoA}}(i) \right) \times \frac{1}{R} \tag{4.20}$$

In Fig. 4.4, a sample of a normal PCG signal from the test database is presented. The following sub-figures illustrate the distribution of different parameters and the likelihood of the critical cut-off intensities. The calculated x_{uc} and x_{lc} values are indicated in Fig. 4.4 (b) and (c) by star symbols respectively.

4.5 Shannon entropy and Shannon energy based mode selection (SE2MS)

In this Section, we present an alternate method of amplitude moderation using either Shannon entropy or Shannon energy, whichever is suitable for a given PCG signal. This mode selection algorithm involves identifying a PCG signal whether it is clean or noisy. For a clean signal (noise free), the algorithm will select Shannon entropy to extract the envelope feature. Otherwise, Shannon energy is selected. This algorithm is motivated from the fact that a majority of the durations of a heart cycle are the silent systolic and diastolic intervals. As a result, the maximum number of signal samples will have small intensity values close to baseline. In a noisy signal, the systolic and the diastolic segments will be affected by noise/ murmurs. Depending on the level of noise intensities and the nature of noise, the number of signal samples at different intensity bin will vary.

To test this assumption, histogram analysis of signal intensities at different intensity levels is performed. Twenty bin histogram with a bin size of 5% of absolute maximum amplitude ($\max(|\mathbf{x}|)$) is considered. The first bin is approximated as the signal intensity level near the isoelectric line representing silent sound. Since the algorithm has to decide from either of the two, Shannon entropy for noise-free signal or Shannon energy for noisy signal, we consider

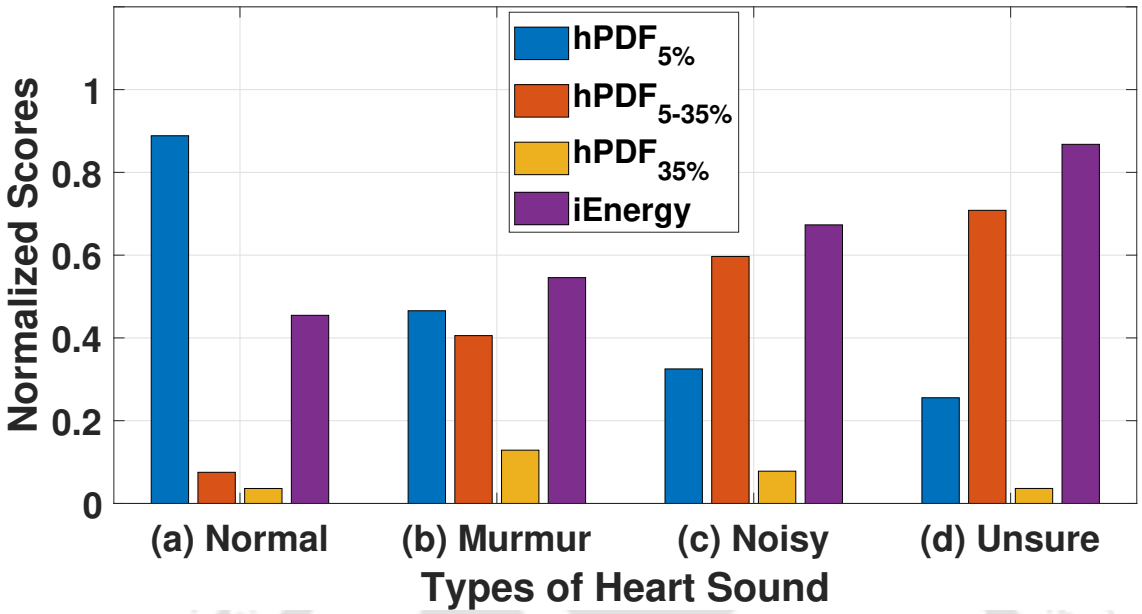

Figure 4.5: Histogram probability density function (hPDF). In panels, (a) clean normal heart sound, (b) heart sound with murmurs and clean silent systole or diastole, (c) noisy heart sound (consistent rumbling noise), and (d) unsure or poor quality signal in which FHS cannot be recognized.

the lower cut-off amplitude of the Shannon energy as the benchmark above which the signal intensities will be mostly of heart sound origin. The signal amplitude (x_{lc}) corresponding to the lower cut-off of Shannon energy transformation is approximately 0.35 or 35% of $\max(|\mathbf{x}|)$. Intermediate intensities ranging between 5% and 35% of $\max(|\mathbf{x}|)$ represent the remaining low-intensity signals that transit to FHS. If noise is introduced, this particular range of signal intensities is affected. From this assumption, the following histogram probability density function (hPDF) and the intermediate signal energy ($\mathrm{iEnergy}$) are calculated.

$$\left.\begin{aligned} \mathrm{hPDF}_{5\%} &= P\left[|\mathbf{x}| \leq 0.05\right] \\ \mathrm{hPDF}_{5-35\%} &= P\left[0.05 < |\mathbf{x}| \leq 0.35\right] \\ \mathrm{hPDF}_{35\%} &= P\left[|\mathbf{x}| > 0.35\right] \end{aligned}\right\} \tag{4.21}$$

$$\mathrm{iEnergy} = \frac{\sum_{n=1}^{N}\{|x^2(n)| : |x(n)| <= 0.35\}}{\sum_{n=1}^{N}|x^2(n)|} \tag{4.22}$$

The purpose of $\mathrm{iEnergy}$ is for measuring the dominance of noise signals over the louder heart sound signals. The high likelihood score of $\mathrm{iEnergy}$ means that the lower intensity noise

signal energy is dominant over the heart sound signal energy. The other interpretation is that the signal-to-noise ratio (SNR) of a PCG signal is low. Intuitively a threshold of 0.8 or 80% of total energy is taken to determine whether FHS envelope is discriminating or should it be categorized as undiscriminating/unsure.

Fig. 4.5 presents an example of hPDF distribution for different types of heart sound recordings. Here, a high value of $\mathrm{hPDF}_{5\%}$ implies that the PCG has relatively clean silent intervals. The $\mathrm{hPDF}_{5-35\%}$ estimates the portion of signal samples that have lower medium intensities. The dominant score of this measure would mean that the signal is noisy. The last measure defined by $\mathrm{hPDF}_{35\%}$ denotes the remaining signal samples that may be categorized as loud sound representing FHS. The $\mathrm{iEnergy}$ also provides similar implication as that of $\mathrm{hPDF}_{5-35\%}$. The difference is that it ensures how much signal information is corrupted by the intermediate signal intensities. This information is quantified as a density function of signal energy. A high value of $\mathrm{iEnergy}$ suggests that most of the signal information are within the intensity range of $5-35\%$. Alternately, this also indicates that there is less information in the higher signal intensities. For an automatic selection of operational modes between Shannon entropy and Shannon energy, the following thumb rule based on hPDFs and $\mathrm{iEnergy}$ is set up, as shown in the Algorithm 3.

Algorithm 3 Algorithm for Shannon entropy/Shannon energy mode selection (SE2MS).

$\mathrm{Shannon\_entropy} = @(x) - |x| \log(|x|)$;
$\mathrm{Shannon\_energy} = @(x) - x^2 \log(x^2)$;
if $\{\mathrm{hPDF}_{5\%} \geq \max(\mathrm{hPDF}_{5-35\%}, \mathrm{hPDF}_{35\%})\}$
  $\mathrm{SE2MS\_E}(n) = \mathrm{Shannon\_entropy}(x(n)); \quad n = 1 : N$
else if $\{\mathrm{hPDF}_{5-35\%} \geq \max(\mathrm{hPDF}_{5\%}, \mathrm{hPDF}_{35\%}) \quad \& \quad \mathrm{iEnergy} < 0.8\}$
  $\mathrm{SE2MS\_E}(n) = \mathrm{Shannon\_energy}(x(n)); \quad n = 1 : N$
else
  DISPLAY 'The signal is uncertain'.
end

4.6 Evaluation process

The evaluation of the proposed algorithm is carried out using the heart sound database available in PhysioNet/Computing in Cardiology Challenge 2016 [4, 18, 71]. It is discussed in Chapter 2 Section 2.1. The augmented database for evaluation of HSS algorithms is available for downloading at [72]. It contains 792 PCG recordings of 135 subjects, sampled at 1 kHz. Of those, 386 PCG recordings are from healthy control subjects, and the remaining 406 signals are from patients suffering from arrhythmia, valvular septal defect (VSD), and other cardiac-related diseases. The database provides the ground-truth label of S1 and S2 sounds in the PCG as the location of R-peak and the end-T-wave of the corresponding ECG signal, sampled at 50 Hz. It also contains a list of binary diagnosis indicating the PCG signal is normal (binary diagnosis = 0) or abnormal (binary diagnosis = 1).

Figure 4.6: Block diagram of evaluation process with modules involved in heart sound segmentation (HSS).

To evaluate the proposed heart sound enhancement methods, the envelope features extracted using these methods are used for segmentation of the heart sound, and their performances are compared. The evaluation process consists of three major modules as shown in the block diagram, Fig. 4.6.

Preprocessing

In this module, the PCG signal is processed to suppress as much noise as it is possible. One attractive method is using the total variation filter (TVF) [56, 105]. In Chapter 3 Section 3.2, a method is proposed, known as dual filtering, that uses the combined effect of the band-pass filter (BPF) and TVF. First, the effect of high-frequency noise is suppressed by passing the PCG signal through a first-order Butterworth band-pass filter within a frequency range of

20 Hz and 150 Hz. This may still leave some traces of noise in the signal. Therefore, the signal is further processed with the TVF filter to smooth out the remaining low-intensity noise signals. The same method is used in this module for denoising a PCG signal.

Envelope extraction

The filtered signal is amplitude moderated using the proposed methods. This will emphasize all the high-intensity sound resulting in uniform peak intensities. From the moderated waveform, the envelope is extracted by the homomorphic envelogram method [106]. Taking this envelope feature and the hidden semi Markov model (HSMM) classifier, the PCG signals are segmented into its four major components; S1, systole, S2, and diastole.

Hidden semi Markov model (HSMM) classifier

HSMM is a statistical model with double stochastic process of the observable sequence $O = \{O_1, O_2, ..., O_t, ..., O_T\}$ and the hidden state sequence $Q = \{q_1, q_2, ..., q_t, ..., q_T\}$. It estimates the single best state sequence Q^* that bears maximum likelihood of the given observation sequence O and the model $\hat{\Lambda}$.

$$Q^* = \arg \max_{Q} P(Q|O, \hat{\Lambda}) \tag{4.23}$$

The HSMM model $\hat{\Lambda} = \{A, B.\pi, p\}$ has four crucial parameters:

(i) $A = \{a_{i,j}\}$ defines the transition probability from state $q_{t-1} = i$ to $q_t = j$. Assuming the heart sound components to occur in a fix state sequence of $\mathrm{S1} \rightarrow \mathrm{systole(Sys)} \rightarrow \mathrm{S2} \rightarrow \mathrm{diastole(Dia)} \rightarrow \mathrm{next\ heart\ cycle}$, its value is preset as under:

$$A = \begin{array}{c} \begin{matrix} \text{S1} & \text{Sys} & \text{S2} & \text{Dia} \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \text{i/j} \\ \text{S1} \\ \text{Sys} \\ \text{S2} \\ \text{Dia} \end{matrix}$$

(ii) $B = \{b_j(O_t)\} = P\left(O_t|q_t = j\right)$ is an emission probability of observing O_t at state j . In [5], the Bayes' rule and logistic regression are used to derive the emission probability:

$$P\left(O_t|q_t = j\right) = \frac{P\left(q_t = j|O_t\right) \times P(O_t)}{P(j)} \tag{4.24}$$

$$P\left(q_t = j|O_t\right) = \sigma\left(w'O_t\right) \tag{4.25}$$

where w is the model weight and $\sigma(x) = \frac{1}{1+e^{-x}}$ is the logistic function. The $P(O_t)$ is calculated as a multivariate normal distribution from the features of the entire training set. Every state of the heart sound has equal chance of occurrence. Therefore, $P(j) = \pi_j = \frac{1}{4}$

(iii) $\pi = \{\pi_i\} = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]$ is the initial state probability.

(iv) In this thesis work in Chapter 5, a multi-mode diastolic duration model $p_{\text{Dia}}(d)$ has been introduced to effectively adjust the derived state sequence with the variability of heart cycle duration. It is a cascaded Gaussian distribution of all possible duration model $(c_i, \varrho_{c_i}, w_i)$.

$$p_{\text{Dia}}(d) = \sum_{i=1}^{n_c} w_i \times \mathcal{N}(c_i, \varrho_{c_i}^2, d). \tag{4.26}$$

where c_i, ϱ_{c_i} and w_i are the respective centroid, the standard deviation and the weight of each cluster obtained from hierarchical agglomerative clustering of all possible diastolic duration estimated from each PCG signal. The maximum length 'd' of the duration

probability is one heart cycle duration [5, 16, 105].

Evaluation metrics

The performance of the proposed LFAM based envelope ($\mathrm{LFAM\_E}$) and SE2MS based envelope ($\mathrm{SE2MS\_E}$) are compared with the envelopes extracted using conventional homomorphic filter ($\mathrm{Homo\_E}$), Shannon entropy ($\mathrm{SEnt\_E}$), and Shannon energy ($\mathrm{SEng\_E}$). Each of these methods is tested against normal and abnormal PCG signals. Acquisition of heart sound signal is not free of aberrations and it may be affected by artifacts and ambient noise. Therefore, several experiments are carried out on noisy PCG signals which are generated at different SNR values (10 dB, 5 dB, 0 dB, -5 dB) by adding white Gaussian noise (AWGN) and real-time ambient noise (RTAN).

The performance of amplitude moderation is evaluated by calculating the mean and standard deviation of the envelope peaks of S1 sound, S2 sound and along with the systolic and diastolic intervals for a given PCG signal. These scores are expected to bear consistent mean values with a minimal variation on an excellent amplitude moderated envelope.

To determine whether the proposed envelope extraction method improves the discrimination of FHS, the LFAM envelope is tested for FHS segmentation using HSMM classifier [105]. Among the other HSS methods in the literature [5, 6, 63, 68, 105], HSMM works well with a small dataset and on short-duration PCG signals. It gives reasonably good segmentation using a small feature dictionary.

The performance of the system generated delineation, and the classification of heart sound is scored using the following standard metrics: Sensitivity (Se), Positive prediction (P_+), and F-score (F_1). A tolerance of $\pm$ 40 ms is permitted for onset and offset of events to match the ground truth [6, 105].

4.7 Results and discussions

In above Sections, we have introduced two different heart sound envelope enhancement algorithms, namely, LFAM based envelope ($\mathrm{LFAM\_E}$) extraction and SE2MS based envelope

(SE2MS_E) extraction . In the Fig. 4.7 the amplitude moderated signals with their envelopes are shown using LFAM and SE2MS methods.

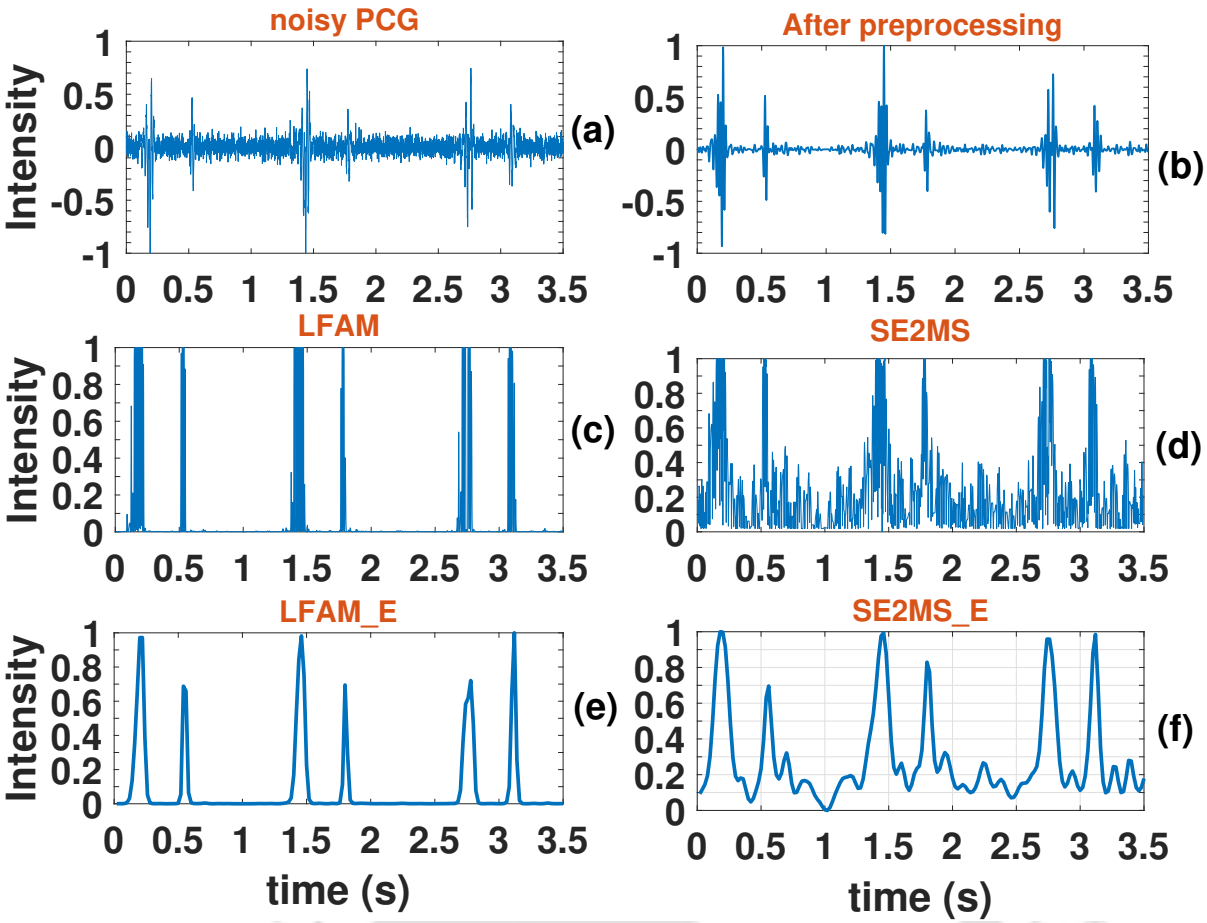

Figure 4.7: LFAM based envelope (LFAM_E) and SE2MS based envelope (SE2MS_E) extraction. In panels, (a) a noisy PCG signal, (b) output after preprocessing, (c) amplitude moderation LFAM, (d) amplitude moderation using SE2MS, (e) the resulting envelopes using LFAM and (f) the resulting envelope using SE2MS.

The Fig. 4.7 (a) and (b) show the noisy PCG and preprocessed PCG signals. The amplitude moderations using proposed methods, LFAM and SE2MS, are shown in Fig. 4.7 (c) and (d) respectively. The Fig. 4.7 (e) and (f) give the resulting envelopes extraction using proposed LFAM and SE2MS based methods. Both the methods perform better for heart sound envelope extraction and preserve the FHS peaks (S1 and S2, Fig. 4.7 (e) and (f)).

4.7.1 Signal dependency

Estimation of the parameters from signal itself improves the LFAM method to enhance the heart sound envelope. Fig. 4.8 and Fig. 4.9 illustrate the results of LFAM ($\sigma_{(x_{\mathrm{lc}},x_{\mathrm{uc}})}$) of heart sound signals that have different S1, S2 and noise intensities and taking different critical x_{lc} and x_{uc} values. The parameters of $\sigma_{(0.03,0.08)}$ in Fig. 4.8 and $\sigma_{(0.08,0.18)}$ in Fig. 4.9 are calculated from the signals. For the data set with lesser noise content in the signal (Fig. 4.8), the value of $\sigma_{\alpha,\beta}$ is lower whereas for signal with higher noise content (Fig. 4.9) the value of $\sigma_{\alpha,\beta}$ is higher.

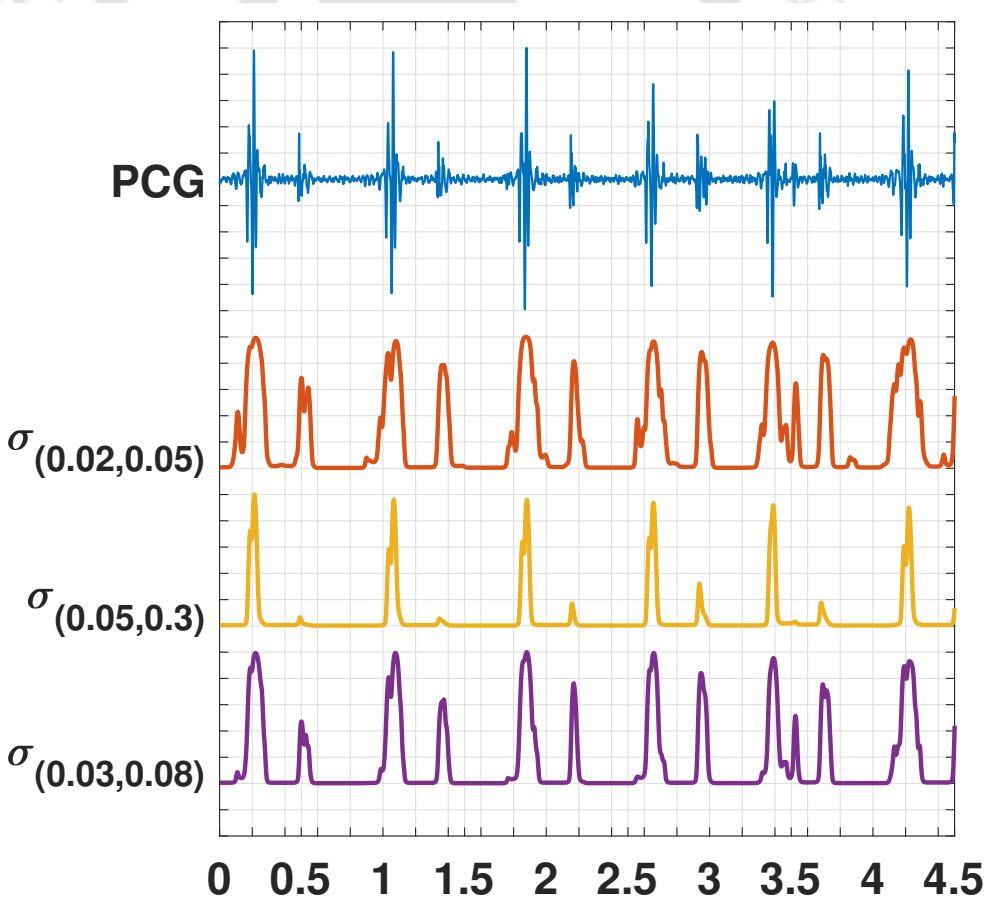

Figure 4.8: A PCG signal $\mathrm{example\_audio\_data}\,\{1,5\}$ of $\mathrm{example\_data}$ and resulting envelopes using $\sigma_{(x_{\mathrm{lc}},x_{\mathrm{uc}})}$ from different x_{lc} and x_{uc} values.

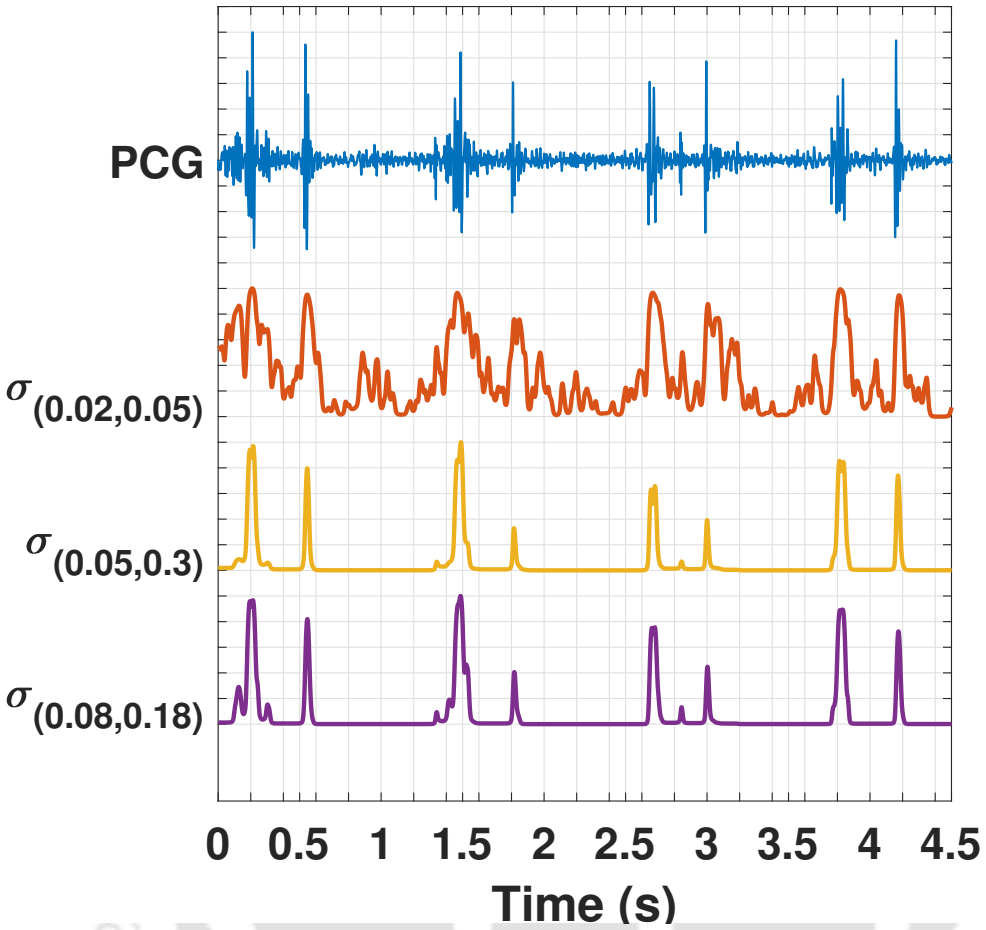

Figure 4.9: A PCG signal $\mathrm{example\_audio\_data}\,\{1, 9\}$ of $\mathrm{example\_data}$ and resulting envelopes from different x_{lc} and x_{uc} values.

From the figures (Fig. 4.8 and Fig. 4.9), it is clear that small values of x_{lc} and x_{uc} are suitable to enhance the heart sound envelope under the minimal or no interference of noise. For a noisy signal, x_{lc} must adjust with the nature of noise and SNR level. Ideally, in a PCG recording, the intensities of S1 and S2 are always higher than the noise signals. Therefore, 20 or 30 percent intensity is a good threshold to segregate the loudest possible noise from heart sound signals. In practical applications, some recordings may have medium to very low-intensity S2 sounds. This makes it challenging to generalize the parameter value for the whole dataset. It is also true when determining the value of x_{uc}. In LFAM, x_{uc} represents the intensity level above which the signal will have uniform intensities. Therefore to obtain the desired uniform envelope peaks of both S1 and S2 sound for the whole dataset, the

parameters of LFAM cannot be predefined but have to be estimated from the individual signal itself. Our proposed method is the closest signal-dependent approach to find these values.

4.7.2 Comparison of envelope peaks

The envelopes are extracted using the conventional homomorphic filter (Homo_E), Shannon entropy (SEnt_E), Shannon energy (SEng_E), and the proposed SE2MS_E and LFAM_E methods in Fig. 4.10 and Fig. 4.11.

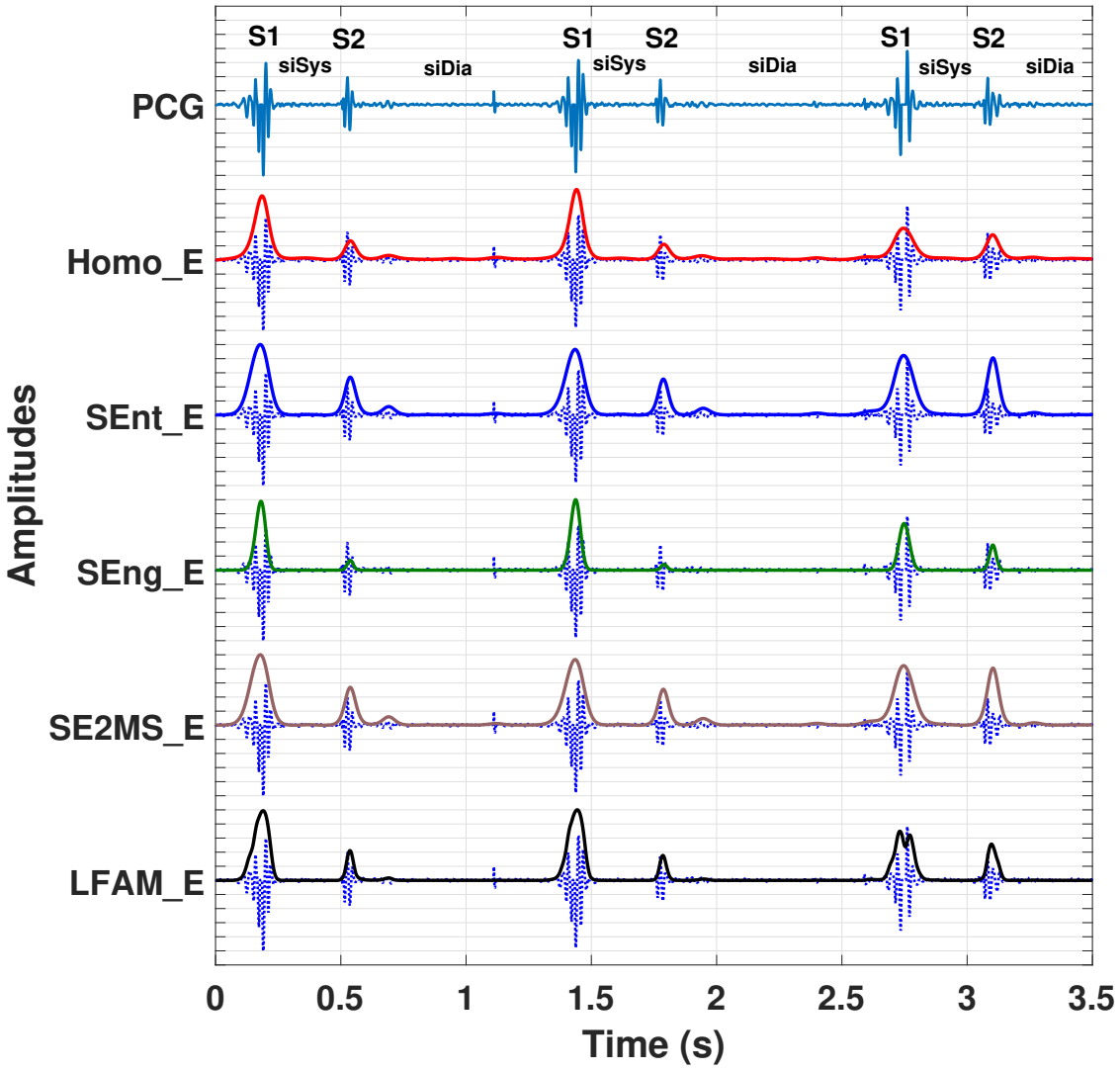

Figure 4.10: Shows envelopes of a noise-free heart sound signals using different envelogram methods.

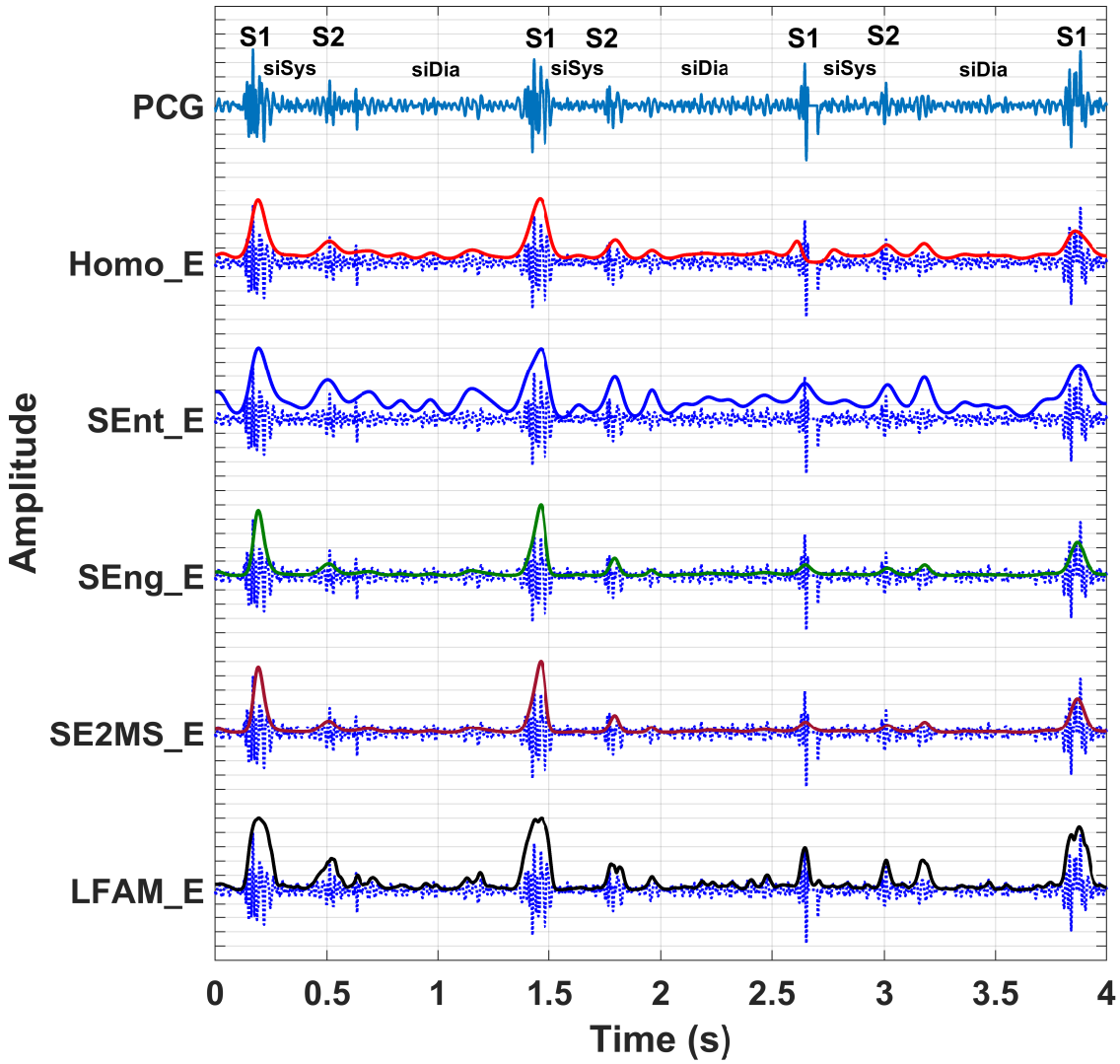

Figure 4.11: Shows envelopes of a noisy heart sound signals using different envelogram methods.

The PCG signal is marked with S1, S2, silent systole zone (siSys) and silent diastole zone (siDia) in both the figures. For the noise free heart sound signal (Fig. 4.10), both the proposed methods, SE2MS and LFAM, give better results in extracting envelopes of heart sound components (S1 and S2). In case of noisy signal (Fig. 4.11), the LFAM based envelope extraction gives better result.

The mean and variance of the envelope peak intensities of the four major heart sound components, S1, S2, silent systole zone (siSys) and silent diastole zone (siDia), are shown in Table 4.1 and 4.2.

Table 4.1: Envelope peak intensities and variance of different heart sound components extracted using proposed methods and conventional methods against normal heart sound signals.

Envelope	**S1**	**siSys**	**S2**	**siDia**
Homo_E	0.63 ± 0.20	0.07 ± 0.05	0.38 ± 0.20	0.07 ± 0.04
SEnt_E	0.82 ± 0.12	0.23 ± 0.08	0.69 ± 0.16	0.24 ± 0.08
SEng_E	0.69 ± 0.21	0.09 ± 0.04	0.49 ± 0.24	0.10 ± 0.05
SE2MS_E	0.77 ± 0.19	0.16 ± 0.08	0.61 ± 0.24	0.16 ± 0.08
LFAM_E	0.93 ± 0.08	0.04 ± 0.07	0.83 ± 0.17	0.04 ± 0.07

Table 4.1, gives the mean peak intensity values of heart sound components with variance for all the methods in case of a normal heart sound signal. In case of FHS (S1 and S2) a higher mean intensity value is desirable whereas for siSys and siDia, a lower mean intensity value is better. The conventional Homo_E gives lower peak intensity values 0.63 for S1 and 0.38 for S2 whereas the proposed LFAM_E yields higher peak intensity values 0.93 for S1 and 0.83 for S2.

Table 4.2: Envelope peak intensities and variance of different heart sound components extracted using proposed methods and conventional methods against abnormal heart sound signals.

Envelope	**S1**	**siSys**	**S2**	**siDia**
Homo_E	0.63 ± 0.21	0.10 ± 0.07	0.37 ± 0.21	0.08 ± 0.05
SEnt_E	0.82 ± 0.14	0.26 ± 0.09	0.66 ± 0.19	0.25 ± 0.09
SEng_E	0.68 ± 0.23	0.10 ± 0.05	0.44 ± 0.26	0.10 ± 0.05
SE2MS_E	0.74 ± 0.22	0.16 ± 0.08	0.54 ± 0.24	0.15 ± 0.08
LFAM_E	0.92 ± 0.10	0.07 ± 0.09	0.77 ± 0.24	0.05 ± 0.07

In case of a abnormal heart sound signal, mean peak intensity values of heart sound components with variance are shown in Table 4.2. The conventional Homo_E gives lower

peak intensity values 0.63 ± 0.21 for S1 and 0.37 ± 0.21 for S2 whereas the proposed $\mathrm{LFAM\_E}$ yields higher peak intensity values 0.92 ± 0.10 for S1 and 0.77 ± 0.24 for S2. The mean peak intensity values indicate the superior performance of our proposed method for both normal and abnormal hear sound recordings.

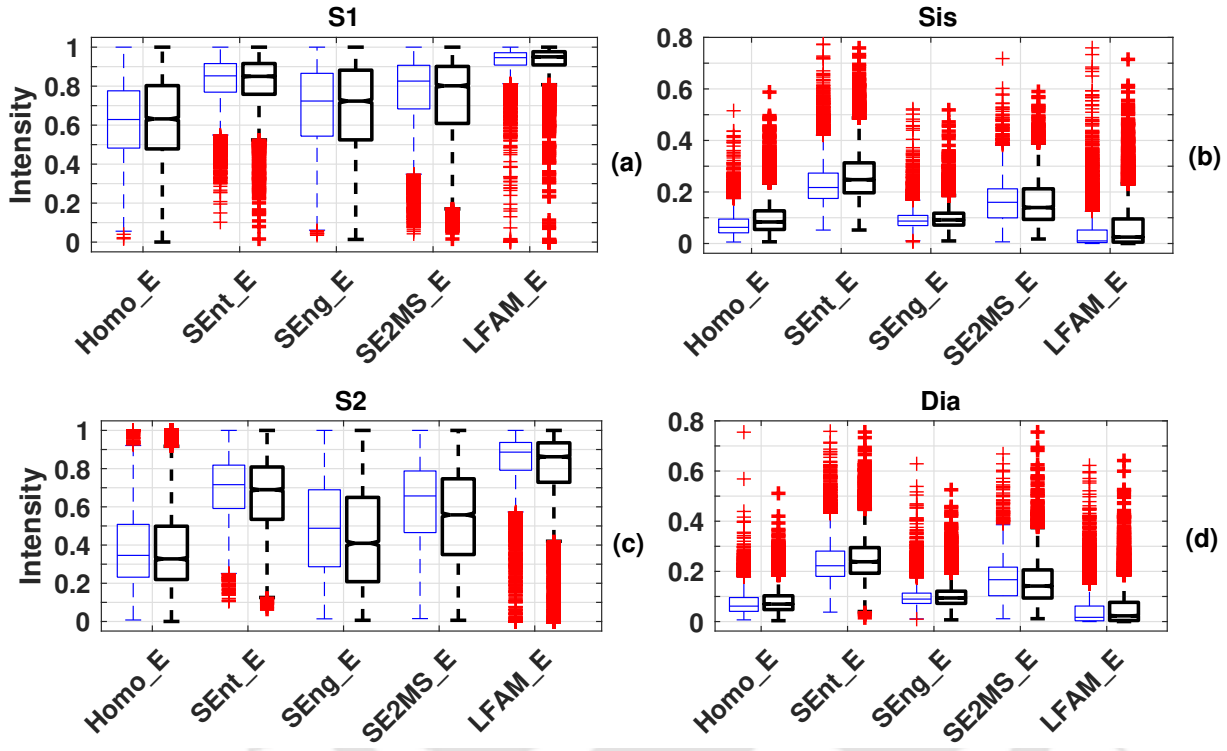

Figure 4.12: The intensity distribution of envelope peaks for (a) S1 sounds, (b) systole intervals, (c) S2 sounds, and (d) diastole intervals, respectively. The envelopes are extracted using the conventional homomorphic filter ($\mathrm{Homo\_E}$), Shannon entropy ($\mathrm{SEnt\_E}$), Shannon energy ($\mathrm{SEng\_E}$), and the proposed $\mathrm{SE2MS\_E}$ and $\mathrm{LFAM\_E}$ methods against normal (blue, thin boxes) and abnormal (thick, black boxes) heart sound signals.

The box plot of the corresponding moderated amplitudes are shown in Fig. 4.12. Table 4.2 (blue, thin boxes in Fig. 4.12) shows the evidence of the moderated amplitude for normal/healthy control heart sound signals. Amongst the conventional methods, Shannon Entropy provides the best emphasis on S1 and S2 sound. However, the results of siSys and siDia clearly show that these methods do not effectively suppress the low-intensity noise. Instead, the noise signals are equally emphasized. This is visible in Fig. 4.10, and Fig. 4.11. On the other hand, the internal algorithm of SE2MS selects the suitable moderator, either Shannon entropy or Shannon energy, based on whether the signal is clean or noisy. Therefore,

there is not much improvement over the others. In the case of LFAM, the emphasis given to S1 and S2 sound are significant. Another improvement is in extenuating low amplitude noise signals in siSys and siDia intervals. The resulting moderated waveform yields distinct discrimination between envelope peaks-and-valley distribution of audible heart sound from noisy/silent systole and diastole intervals.

For analysis of noisy and abnormal heart sounds, the Shannon energy envelope is the benchmark method to suppress the noise signal amplitude in the expected silent sound intervals of the heart sound. As observed in Table 4.3 (thick, black boxes in Fig. 4.12), the proposed method shows much enhancement of FHS. Both SE2MS and LFAM improve the difference in intensity for inter heart sound components; distinctly observed in the case of LFAM methods.

4.7.3 Evaluation of proposed methods for heart sound segmentation

To compare the performance of the existing and the proposed moderated envelope features for HSS, the hidden semi Markov model (HSMM) is used as a classifier. Table 4.3 shows the results of HSS at different SNR levels of additive AWGN and clinical noise. For PGC recordings with additive noise at large SNR, the scores are comparable for all the envelogram methods. The results are because HSMM classifier derives the state sequence based on the probabilistic approach using maximum likelihood criterion. When the signals are relatively clean (less interference from noise), the envelope peak intensities of S1 and S2 sounds are much higher than the silent sound intervals. Therefore the estimated likelihood of S1 and S2 will have maximum value and generates correct labels against their ground truth. The evaluation results are obtained over 30 iterations. In every iteration cycle, the dataset was randomly split into two halves, each with an equal proportion of normal and abnormal heart sounds. One half will be used for training while the other for testing. The gross performance scores of the five envelope extraction methods are presented below.

For noisy PCG recordings with low SNR values, 0 dB and -5 dB, the proposed LFAM appears to yield better segmentation accuracy (shown in bold face numericals in Table

Table 4.3: Evaluation of the Conventional Amplitude Moderation methods against the proposed methods using HSMM classifier for heart sound segmentation, and across 30 iterations.

Envelope	SNR in dB			
	10	**5**	**0**	**-5**
Homo_E	$\mathbf{96.59 \pm 0.25}$ 96.85 ± 0.29	$\mathbf{95.78 \pm 0.37}$ 96.40 ± 0.35	$\mathbf{95.12 \pm 0.80}$ 95.64 ± 0.78	$\mathbf{92.31 \pm 2.17}$ 91.48 ± 2.25
SEnt_E	$\mathbf{75.27 \pm 0.18}$ 97.32 ± 0.14	$\mathbf{97.25 \pm 0.23}$ 97.28 ± 0.27	$\mathbf{96.71 \pm 0.52}$ 96.54 ± 0.64	$\mathbf{92.15 \pm 2.12}$ 93.84 ± 2.07
SEng_E	$\mathbf{97.26 \pm 0.20}$ 97.23 ± 0.16	$\mathbf{96.87 \pm 0.21}$ 97.11 ± 0.24	$\mathbf{96.45 \pm 0.51}$ 96.56 ± 0.53	$\mathbf{94.10 \pm 1.83}$ 94.46 ± 2.03
SE2MS_E	$\mathbf{97.38 \pm 0.15}$ 97.35 ± 0.08	$\mathbf{97.25 \pm 0.27}$ 97.21 ± 0.20	$\mathbf{96.78 \pm 0.31}$ 97.25 ± 0.14	$\mathbf{94.29 \pm 2.01}$ 95.15 ± 1.61
LFAM_E	$\mathbf{97.45 \pm 0.15}$ 97.73 ± 0.11	$\mathbf{97.38 \pm 0.17}$ 97.50 ± 0.09	$\mathbf{97.19 \pm 0.36}$ 97.23 ± 0.21	$\mathbf{95.13 \pm 1.03}$ 95.36 ± 0.80

It shows only F-score or F_1 (%). Those in bold are results from the signal affected by AWGN noise. The second scores are measured under the influence of RTAN.

4.3). This improvement is partially due to the preprocessing step. The dual filtering process using TVF helps to estimate the approximate heart sound signal by smoothing the high-frequency, low amplitude noise signals. This process will leave us with some high-intensity noise signal which is responsible for a noisy envelope. As explained earlier, the Shannon entropy will emphasize such noise signals making the envelope noisy. The homomorphic and Shannon energy envelope are less affected by noise, but the low amplitude heart sounds may disappear with noise signals. SE2MS will suffer the same losses. The proposed LFAM envelope may appear noisy for some signals. The purpose of this method lies in the fact that it emphasizes any high amplitude signals from the surrounding and suppresses the average lower amplitude distribution as silent sound intervals. The resulting envelope significantly improves the estimated likelihood of the S1 and S2 sounds in the classification module.

The proposed logistic function amplitude moderation based envelope extraction method is versatile towards different SNR levesl of noise in PCG or heart sound recordings. Estimation

of critical upper and lower cut-off amplitudes from the signal itself improves the mapping of the sigmoid curve to the desired segregation of heart sound signals from the weak noise signals. This method is an improvement over the Shannon energy and Shannon entropy, by including higher intensity signals (approaching absolute maximum value) for extraction of envelope features. This incorporates the accent of S1 and S2 as perceived (loud) sounds in a heart cycle.

The other advantage of LFAM is in generating relatively uniform envelope peaks with a minimal difference, as shown in Fig. 4.12. Proper calibration of trailing S-curve also improves the suppression of weaker noise signal in PCG or heart sound recordings. This helps in enhancing the difference between inter heart sound components (S1, systole, S2, and diastole). The resulting envelope is less affected by noise and yields better segregation of audible heart sound from expected silent intervals, illustrated in Figs. 4.8, 4.9, 4.10, and 4.11.

4.8 Summary

This Chapter presents a new method to extract the envelope of the fundamental heart sound (S1 and S2) using the logistic function. The sigmoid characteristic of the logistic function is incorporated to segregate S1 and S2 signal intensities from silent or noise interfered systolic and diastolic intervals in a heart sound cycle. This signal intensity transformation brings uniformity to the envelope peaks of S1 and S2 sounds by inclining the transform intensity distribution towards the upper asymptote of the sigmoid curve. The method also helps in subduing noise in silent systole and diastole intervals of heart sound sequence by mapping the low amplitude noise intensities to the asymptotic sigmoid tail. The proposed logistic function-based amplitude moderation (LFAM) envelogram method involves finding the critical upper amplitude (x_{uc}) and the critical lower amplitude (x_{lc}) from the heart sound signal. These critical amplitude values are used to determine the scaling and shifting parameters of the logistic function, which is used in LFAM to generate the moderated envelope of the

signal. Their values are regressively obtained from the signal by histogram analysis of intensity distribution using a heuristic technique. That way, the derived parameters are data-dependent. This enable the transformation function, defined by the logistic function ($\sigma_{(x_{\mathrm{lc}},x_{\mathrm{uc}})}$), to adjust the signal intensity distribution automatically in the most suitable way to improve the envelope feature. The performance is carefully assessed using heart sound data recorded under controlled and clinical environments and at different SNR levels. The proposed LFAM based envelope feature proof to enhance the heart sound envelope peaks better than other existing envelope features. When evaluated for heart sound segmentation using HSMM, the LFAM envelope feature significantly improves segmentation accuracy, especially at a low signal-to-noise ratio. The best average F1 score achieved is 97.73%.

5

Heart sound segmentation

Contents

In the previous Chapter, two novel algorithms are discussed to extract and enhance the envelope of heart sound. The enhancement of FHS envelope may help in segmentation of heart sound more accurately. In this chapter, a multi-modal duration distribution is introduced as an extension of the hidden semi-Markov model (HSMM) based heart sound segmentation (HSS) algorithm. In this model the derived state will adjust itself with all possible variation of heart cycle duration (HCD). In the proposed multi-mode duration model, each mode represented by weighted density distribution acts as the center of a search space where the actual duration of the heart sound component is expected. To achieve this, the HCDs are estimated at multiple instances of a PCG to get all possible duration values. Then the nearest HCD values are clubbed together into larger clusters so that each centroid represents the consolidated reference point. The cluster size should not hold two heart sound candidates within its distribution. Finally, the Gaussian density functions of these centroids are cascaded into one to form the proposed state duration model of a subject.

The introduction of duration-dependent classifiers for heart sound segmentation is largely motivated by the traditional auditory perceptual method. During auscultation, experts and physicians listen to the heart sound and intuitively identify S1 and S2 based on prior knowledge of the intensity, timber, expected duration of S1, S2, and the difference in the time interval between them. Both S1 and S2 sounds have similar morphology and they are undistinguished. However, their temporal locations in a cardiac cycle are fixed, i.e. (a) a heart cycle consists of S1, silent systole, S2 and silent diastole in sequential order, and (b) the interval between S1 and S2 is always shorter than the interval between S2 and S1, irrespective of the heart rate. These assumptions have been a key to discriminate the FHS in many research works and clearly simplifies the HSS problem. In HSMM, the duration model acts as a guideline to stipulate the approximate duration of each state. This avoids the probable misclassification of states within the expected time interval. The Markov model is designed to be non-ergodic. This sets the derived state sequences to occur in a fixed pattern. The whole setup of HSMM closely meets the required model to develop a heart sound segmentation algorithm.

Gamero and Watrous [60] introduced the HMM for HSS using mel-frequency cepstral coefficients. The doubly stochastic process of this model corresponds well with the current task description, where the observable features of the heart sounds are given but the corresponding states are unknown. There are many attempts to explore the attributes of this algorithm by incorporating different feature extraction methods [61, 62]. The standard HMM has one major drawback. The expected duration of each state is not explicitly included in the model. This poorly suits the identification of heart sound components where duration information is of utmost importance.

Gill *et al.* [62] introduced the time duration information of each heart sound component as an additional feature to model the duration dependent HMM, known as the hidden semi-Markov model (HSMM) which achieved higher segmentation accuracy. Both Schmidt *et al.* [16] and David *et al.* [5] rigorously analyzed the HSMM model on a large dataset, and demonstrated the superiority of this method. The incorporation of the state duration avoids misclassification of states within the expected interval.

HSS using deep recurrent neural network (DRNN) [6] is another state-of-the-art algorithm whose performance is at per with the HSMM algorithm. But, it has an edge over the HSMM in identifying the sound components such as S3 and S4 sounds. It does not require any prior information of state duration during training. The features used in their experiment are 41-bin log magnitude spectrogram (LogMS), 20 coefficients each of mel frequency cepstral coefficients (MFCC), phase residual delta coefficients (Δ) and acceleration coefficients (Δ^2), along with the four envelope features [5]: Homomorphic envelope, Hilbert envelope, wavelet envelope and PSD envelope.

In the following Sections, we will discuss how the duration of four major heart sound components are determined, and the corresponding state duration density functions are modeled. It will also present the stepwise procedures involved in developing the proposed multi-modal diastolic duration model. The proposed modification in HSMM is evaluated for heart sound segmentation using different envelope features, and their results are discussed.

5.1 Modelling of state duration

The HSMM model used for the current HSS is a four-state model. These states represent the four major components of a heart sound, i.e. S1 sound, silent interval of systole (siSys), the S2 sound and the silent interval of diastole (siDia). As reported in literature [5, 16], both S1 and S2 are short duration sounds and they are assumed to be relatively constant for all PCG recordings irrespective of the subjects. The average duration of the S1 and S2 sounds are approximately 122 ms and 92 ms respectively with a permissible tolerance of 22 ms [16]. On the other hand, the heart rate of a subject is not always constant but changes depending on the physical and mental health of the subject. Likewise, the duration of diastole and systole are also variable. These duration values cannot be predetermined and have to be estimated from the signal itself. The standard method to calculate HCD is through autocorrelation analysis of a signal envelope. From the autocorrelation function, the highest peaks between 500 ms and 1800 ms time instances is taken as one HCD [5, 16, 107], as illustrated in Fig. 5.1.

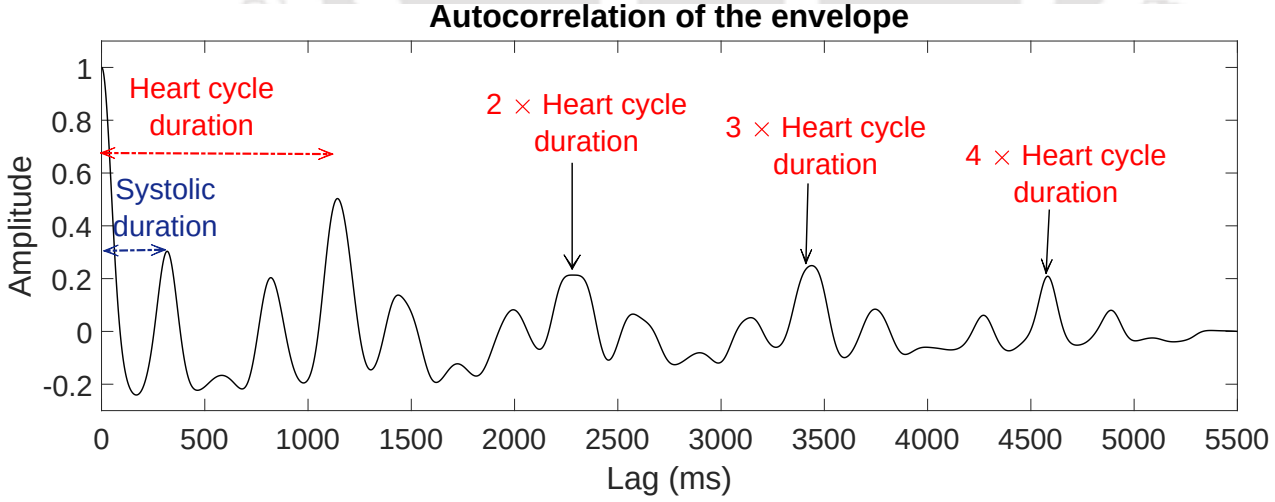

Figure 5.1: Estimating HCD and systolic duration from the autocorrelation function of a PCG envelope.

From our application, the silent systolic duration (d_{siSys}) is defined as the duration from the end of S1 sound to the beginning of the S2 sound. Its mean value and standard deviation are

calculated as:

$$\mu d_{\mathrm{siSys}} = \mu(\mathrm{ed}_{\mathrm{Sys}}) - \mu d_{\mathrm{S1}} \tag{5.1}$$

$$\varrho_{\mathrm{siSys}} = \mathrm{std}(\mathrm{ed}_{\mathrm{Sys}}) \tag{5.2}$$

To estimate all expected duration of these states and measure the degree of variation, the HCD is calculated from every signal segments of 2000 ms duration, each segment starting from the end of the previous heart cycle. The possible end of the heart cycle is extrapolated from the resulting HCD by adding the location of the start and the HCD duration. To validate the HCD, the first and the second highest peak in the autocorrelation function are located. If the ratio of these peak locations, taking the larger value at numerator, is approximately twice i.e. $\in [1.9, 2.1]$, then the minimum value is taken as HCD. Otherwise, the highest peak will be accounted. The average value of all estimated HCD is taken as the ultimate HCD of that subject. Similarly, the estimated duration of systole ($\mathrm{ed}_{\mathrm{Sys}}$) is calculated multiple times in the interval between 200 ms and the minimum of either 550 ms or half of the estimated heart cycle duration. The $\mathrm{ed}_{\mathrm{Sys}}$ corresponds to the time interval from the beginning of S1 sound to the beginning of S2 sound.

Similarly, the silent diastolic duration (d_{siDia}) is defined as the duration from the end of S2 sound to the beginning of the S1 sound. To simplify the calculation, it is taken as the average value of HCD minus sum of the mean value of estimated systolic duration ($\mu\mathrm{ed}_{\mathrm{Sys}}$) and the mean value pf S2 sound duration (μd_{S2}).

$$\mu d_{\mathrm{siDia}} = \mu(\mathrm{HCD}) - (\mu\mathrm{ed}_{\mathrm{Sys}} + \mu d_{\mathrm{S2}}) \tag{5.3}$$

Studies have shown that in a normal subject, few successive heartbeats may have a certain variation of about ± 5 or ± 10 beats/min from its average value. This suggests that a successive HCD may not always have the same value but have some duration difference which could be at most 150 ms [108, 109]. Therefore, it is safe to consider the variation of siDia is within the permissible limit of 100 ms. Schmidt *et al.* [16] have derived ϱ_{siDia} from

siDia through regression analysis:

$$\varrho_{\text{siDia}} = 0.07 \times \mu d_{\text{siDia}} + 6 \text{ ms} \tag{5.4}$$

From the defined mean and standard deviation of each state duration, the duration model is defined as the Gaussian distribution, illustrated in Fig. 5.5.

5.2 Proposed multi-modal diastolic duration distribution

In the presence of anomalies such as deep breathing, sudden physical or emotional stress or pathology, there may be sudden jump in heart rate. Such inconsistency is observed in our working dataset, which shows large difference between minimum HCD and maximum HCD of up to 400 ms in few PCG recordings. In case of siDia, the difference between the minimum ($\text{dmin}_{\text{siDia}}$) and maximum ($\text{dmax}_{\text{siDia}}$) might go as high as 350 ms. The duration distribution with such large deviation will result in poor distribution gradients and may not justify the likelihood of the candidate observation from the other neighboring observations. This complication is observed when the boundary limits given by $\text{dmin}_{\text{siDia}}$ and $\text{dmax}_{\text{siDia}}$ is greater than the systolic duration and both S1 and S2 can be accommodated. To improve the likelihood estimation of the expected state duration, a multi-mode duration model is proposed.

The intuition behind the multi-modal distribution is that each mode will act as a reference location where the actual state may originate. The multiple peak distribution also provides a sharp gradient of the likelihood and improves the discriminability of any two similar observations. Therefore, the mode acts as a local search window with weighted probability that helps estimate the variable state duration more accurately. These attributes make the proposed model more effective than single peak duration model.

In the proposed model, the mode is a Gaussian distribution defined by each cluster centroid of HCD data points. These centroids are selected in such a way that the modal spread may not hold both S1 and S2 sounds together. Therefore, the consecutive centroids should be at least a systolic duration apart. In order to find the best centroid locations, the

HCD data points are first clubbed into denser clusters. Most of the PCG recordings are short duration and the number of HCD points available for analysis is small. For such small data points, which is also one-dimensional information, a simple hierarchical agglomerative clustering based on Ward's method will serve the purpose. The clustering process starts by taking every data point as an individual cluster and then successively merges nearest two clusters to form a new one. The center of new cluster is calculated as [110, 111]

$$c_{i\cup i'} = \frac{\hat{n}_i c_i + \hat{n}_{i'} c_{i'}}{\hat{n}_i + \hat{n}_{i'}} \tag{5.5}$$

where $\hat{n}_i$ and $\hat{n}_{i'}$ are the number of data points in cluster c_i and $c_{i'}$ respectively. The Lance-Williams dissimilarity formula for the new cluster is given by

$$\begin{aligned} d\left(i \cup i', i''\right) = \frac{\hat{n}_i + \hat{n}_{i''}}{\hat{n}_i + \hat{n}_{i'} + \hat{n}_{i''}} d\left(i, i''\right) + \frac{\hat{n}'_i + \hat{n}_{i''}}{\hat{n}_i + \hat{n}_{i'} + \hat{n}_{i''}} d\left(i', i''\right) \\ - \frac{\hat{n}_{i''}}{\hat{n}_i + \hat{n}_{i'} + \hat{n}_{i''}} d\left(i, i'\right) \end{aligned} \tag{5.6}$$

and

$$\hat{n}_{i\cup i'} = \hat{n}_i + \hat{n}_{i'} \tag{5.7}$$

This process will continue until a single cluster is obtain forming a decision tree. It may be terminated at any node based on the required number of clusters or the degree of dissimilarity between clusters. In our work, the number of clusters is decided by limiting the variance of the data points in each cluster such that the distance between the nearest two centroids is always less than the μd_{siSys}. The use of μd_{siSys} as deciding parameter enables the clustering process to satisfy the condition of scale-invariance. It also ensures that no two candidate heart sounds appear within this distribution. The outcome may have only one cluster or more than one cluster depending on the variability of the HCD. For $\hat{n}_c$ (≥ 1) number of cluster centroids, the cascaded Gaussian distribution function for diastolic duration mode is calculated as

$$p(d_{\text{siDia}}) = \sum_{i=1}^{\hat{n}_c} w_i \times \mathcal{N}(c_i, \varrho_{c_i}^2, d_{\text{siDia}}) \tag{5.8}$$

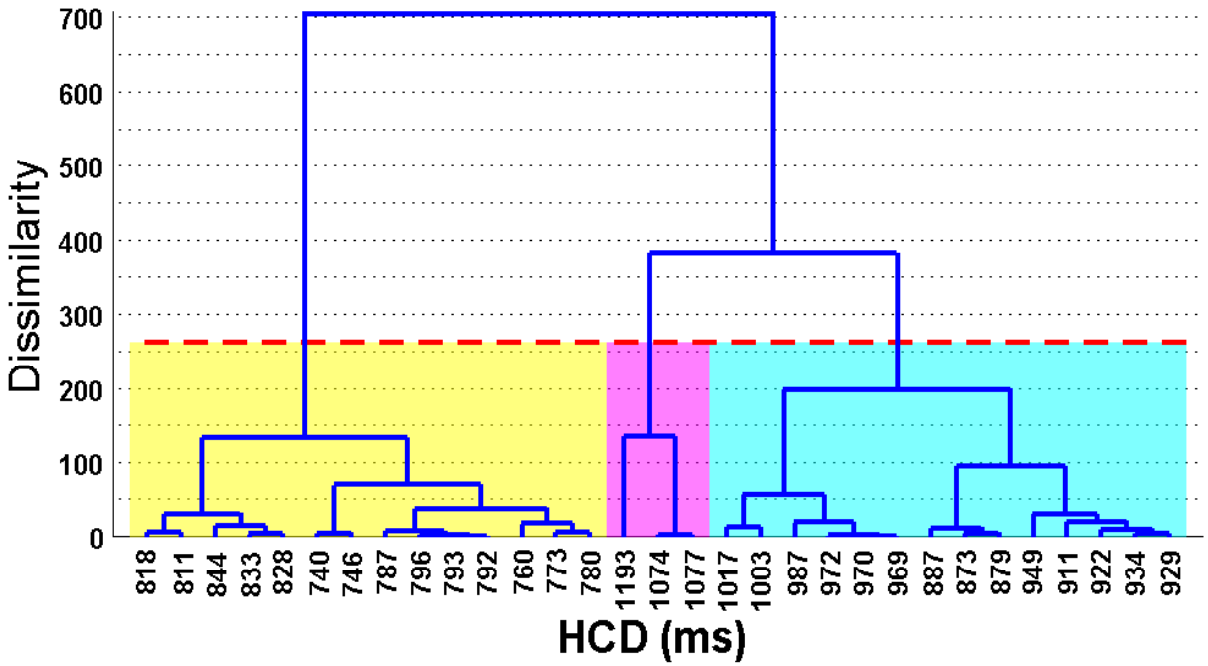

Figure 5.2: The clusters produced at different dissimilarity values are shown in the dendrogram. The data points of HCD is taken from one of the test subject.

where w_i is the weight of each class defined by

$$w_i = \frac{n_i}{\sum\limits_{<n_c>} n_i} \tag{5.9}$$

and ϱ_{c_i} is the standard deviation of each class defined by

$$\varrho_{c_i} = \max\left(\varrho_{c_i}, \varepsilon\right) \tag{5.10}$$

The $\varepsilon = 25$ ms value is used to avoid the problem of singularity when a cluster has only one data point or has very small variation. The $p(d_{\text{siDia}})$ value beyond the range of $\text{dmin}_{\text{siDia}}$ and $\text{dmax}_{\text{siDia}}$ is equated to zero. Fig. 5.2 shows an example of the hierarchical clustering decision tree of one of the test subjects which is terminated when the dissimilarity value is less than $\mu d_{\text{siSys}} = 262$ ms. The corresponding $p(d_{\text{siDia}})$ distribution is shown in Fig. 5.7. This illustrates that the clustering of data points into smaller clusters improves the localization of data by reducing the size of search space.

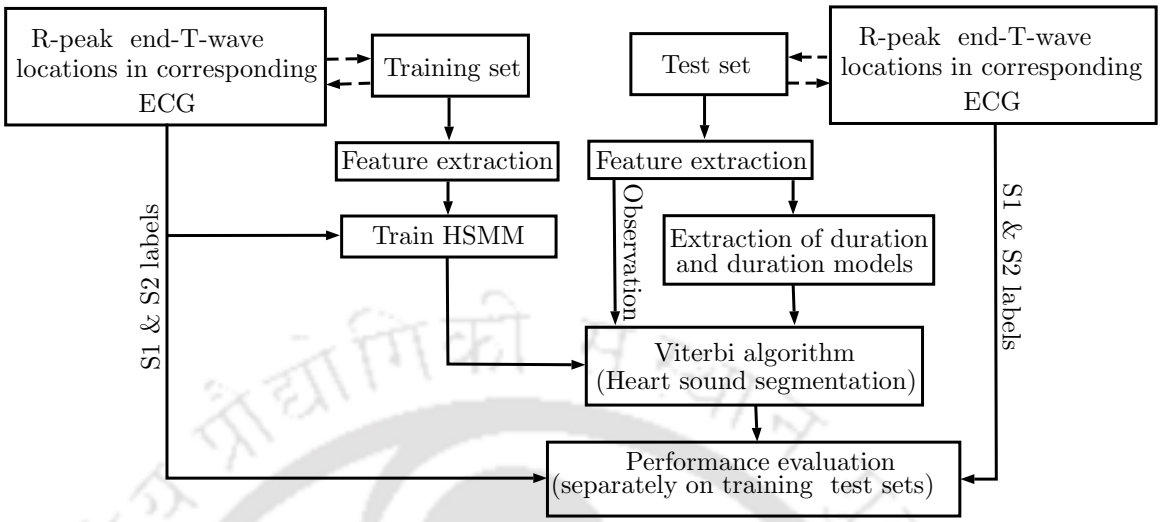

Figure 5.3: Block diagram for HSMM based heart sound segmentation algorithm.

5.3 Evaluation process

The flow diagram of the HSMM based HSS algorithm is illustrated in the Fig. 5.3. In this frame work, the locations of R-peaks and end-T-waves of an ECG signal are used as labels to denote the position of S1 and S2 sounds in the corresponding synchronous PCG signal. This annotation is used only for training and evaluation of segmentation accuracy. The modules involve in this algorithm is discussed in the following section.

5.3.1 Dataset

The detailed description of the dataset used in this experiment is already discussed in Chapter 2 Section 2.1. The dataset has PCG signals with corresponding synchronous ECG signals sampled at 1 kHz and 50 Hz, respectively. To initiate the experiment, the dataset is randomly split into two halves, such that each half contains the same number of recordings from normal and pathological subjects. One-half is used for training and the remaining for testing. Noisy PCG signals are simulated by introducing additive white Gaussian noise (AWGN), and real-time ambient noise (RTAN) at different SNR levels. For the convenience of noise analysis, the RTAN are broadly categorized as (a) air-condition (AC) or ventilation noise, (b) Ambulance siren noise, (c) hospital ambient noise.

5.3.2 Feature extraction

An envelope of PCG signal can impart vital information in the form of envelope peaks rhythmically separated by the depressed segment of the envelope that can easily be correlated with the probable locations of FHS and the silent intervals of systole and diastole based on the duration differences. In clinical practice, such information is intuitively used by a physician or an expert to analyze the heart sound. Frequency analysis also reveals that the fundamental heart sounds have the majority of the signal energy below 150 Hz with a peak at 50 Hz [19]. Therefore, the features used in this work are defined to closely meet the above description. As discussed in [5, 6, 112], four envelope features are extracted: the homomorphic envelope, Hilbert envelope, wavelet envelope, and power spectral density (PSD) envelope.

The homomorphic envelope and Hilbert envelope are signal intensity-based features that are easily affected by the presence of noise and murmurs. Before extracting these features, the PCG signals are preprocessed. In this work, the proposed dual filtering technique is incorporated as a denoising tool before envelope extraction. The combination of BPF and TVF has a better noise reduction capability compared to the conventional low-pass filter [56]. The details of the proposed preprocessing method is discussed in Chapter 3.

From the analysis of PSD of heart sound, it is found that the spectral peak occurs approximately at 50 Hz [5]. This is true for a healthy subject. In pathological condition such as myocardial infarction, cardiomyopathy, and valvular defects, the spectral peak of the S1 and S2 sounds may deviate depending on the elastic property or thickness of myocardium, and the pressure built up in the heart chambers [28]. The exact peak frequencies associated with different pathological cases are not fully studied. Therefore, the peak frequency of the FHS is found from within the bandwidth of 20 Hz and 150 Hz. The final feature is calculated as the PSD of the peak frequency. The Hamming window of width 50 ms with 50% overlap is used for short-term-Fourier-transform of the PCG. This window size is the shortest expected duration of the FHS that can cover its spectral information [5].

All these features are normalized to zero-mean unit variance and later downsampled to

50 Hz in order to minimize the computation cost.

5.3.3 Estimation of parameters

The standard method to train the HSMM parameters ($\pi_i, a_{ij}, b_j(O_t)$) is using the Baum–Welch algorithm [113, 114]. It iteratively estimates the model parameters by maximum likelihood criterion. It makes use of the forward-backward algorithm to compute the statistics for the expectation step without using the information of the true states of observations. But this method may not be feasible for the analysis of PCG signals, which have identical instantaneous observations at different states. That means a likelihood of similar observation sequence may be maximized at multiple states. Therefore without using prior knowledge of true state, the EM algorithm may result in incorrect estimation. In addition to this, the computation involves is time-consuming. Therefore, the parameter values of the model are defined as a set of global parameters applicable to all recordings, and they are determined using statistics from a training dataset where the true states are known.

Considering the fact that the initial state of a PCG signal can be any of the four states, the initial state probability π_i is equated to $\frac{1}{4}$. In defining transition matrix, introducing duration model into HSMM rule out the need of self-transition probability. By assuming that the heart sound components always occur in a fixed order, i.e., $\underbrace{\mathrm{S1} \rightarrow \mathrm{systole(Sys)} \rightarrow S2 \rightarrow \mathrm{diastole(Dia)}}_{\text{one heart cycle}} \rightarrow \mathrm{next\ heart\ cycle}$, there is only one possible transition from each state and its value is rendered unity.

$$A = \begin{array}{c} \begin{matrix} \text{S1} & \text{Sys} & \text{S2} & \text{Dia} \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \text{i/j} \\ \text{S1} \\ \text{Sys} \\ \text{S2} \\ \text{Dia} \end{matrix}$$

5.3.3.1 *Estimation of observation probability*

During the training phase, locations of R-peak and end-T-wave in the corresponding ECG are used as the labels of S1 and S2 sounds in the PCG signal. The duration of S1 sound is considered to begin at the location of R-peak and for a period of mean-S1-duration (μd_{S1}). Similarly, the S2 sound is identified around the location of the end-T-wave that has the maximum S2 sound amplitude. By taking this position as the center and the window length of mean-S1-duration (μd_{S2}), the S2 sound interval in PCG is approximated. The values for μd_{S1} and μd_{S2} are discussed in Section 5.1. The features of each state/component from all the training data are cascaded into one class, forming a feature matrix. Finally, the emission probability '$b_j(O_t)$' is derived from the feature matrix using logistic regression (LR) classifier based on a one-verses-all approach [5].

5.3.4 Testing

For testing, the state duration models were estimated from the individual signal itself. The formulation of existing and proposed duration models are explained in Section 5.1 and 5.2. The performance of these two models are compared by conducting the following experiments.

- experimentation using only homomorphic envelope [16].
- experimentation using homomorphic envelope after proposed dual-filtering (HEoDF).
- experimentation using the homomorphic, Hilbert, PSD, and wavelet envelope features [5].
- experimentation using the HEoDF, Hilbert envelope after dual-filtering, modified PSD, and wavelet envelope features.

The performance is measured based on how accurately the S1 and S2 sounds are detected. The reference locations used for validation of the S1 and S2 sounds are the locations of R-peak and the end of T-wave in the corresponding ECG respectively. These references are the approximate positions to indicate the beginning of the heart sounds. There

is some tolerance in this assumption and it is taken to be within 50 ms as suggested in [5]. If the start of the detected S1 and S2 sounds are located within 50 ms of their respective references then it is considered as truly positive (TP). Otherwise, they are graded as falsely positive (FP). If the heart sounds are not detected at the intended reference position then it is considered as falsely negative (FN). Then, the sensitivity (Se), positive predictivity (P_+) and F_1 score were measured. The observations are repeated multiple times over 30 iterations for both the algorithm taking random training set and test set.

5.4 Results and discussion

The gross performances of the existing and the proposed algorithm with a modified state duration model using different features are shown in Table 5.1. These scores are calculated by summing the total number of truly detected S1 and S2 sounds. Then based on the total TP, FP, and FN count for the whole test dataset, the performance measures are estimated. The final results are the average values from over 30 iterations by randomly selecting training and test set at each iteration. The standard deviation of each score is also shown to illustrate the robustness of the algorithm. The results of the current state-of-the-art algorithm [5] are shown in the 3^{rd} row of Table 5.1, which yields an average F_1 score of 95.40 ± 0.43. The proposed modification, incorporating a multi-modal diastolic duration model, performs better with an average F_1 score of 98.36 ± 0.17.

The proposed method is further evaluated with different additive noise sources, such as (a) additive white Gaussian noise (AWGN), (b) air conditioner (AC) or ventilation noise, (c) ambulance vehicle noise with a siren, and (d) hospital ambient noise. These noise signals are added to clean PCG at different signal-to-noise-ratio (SNR) varying between -5 dB and 10 dB in steps of 5 dB. High SNR values (-5 and 0 dB) are considered to represent noise dominant PCG recordings without proper insulation from the ambient interference. Low SNR values will represent the usual noisy recordings of PCG. The testing is done for each noise type and at different SNR levels. to examine the performance of the proposed dual-filtering process-based

Table 5.1: Comparison of performance scores (%) of existing and proposed methods using various input features and across 30 iterations.

	Features	Classifier	Se	P_+	F_1
*	Homomorphic [16]	LR-HSMM	95.25 ± 0.50	95.85 ± 0.43	95.55 ± 0.45
		Proposed	97.83 ± 0.18	98.24 ± 0.11	98.03 ± 0.13
*	HEoDF	LR-HSMM	95.38 ± 0.41	96.00 ± 0.23	95.69 ± 0.30
		Proposed	98.08 ± 0.19	98.50 ± 0.14	98.29 ± 0.16
*	Hilbert, Homomorphic, PSD, Wavelet [5]	LR-HSMM	95.02 ± 0.51	95.78 ± 0.39	95.40 ± 0.43
		Proposed	97.99 ± 0.24	98.58 ± 0.16	98.28 ± 0.19
*	HEoDF, modified PSD, Hilbert and Wavelet	LR-HSMM	95.32 ± 0.48	96.06 ± 0.54	95.69 ± 0.49
		Proposed	98.28 ± 0.19	98.45 ± 0.16	98.36 ± 0.17

Table 5.2: Results of the proposed algorithm train on 50% of PCG data and tested on the whole dataset with additive noise at various signal-to-noise ratio (SNR) and across 30 iterations.

SNR in dB	Features	Types of noise			
		AWGN	AC	Ambulance	Hospital
-5	a	95.21 ± 0.41	96.65 ± 0.25	93.29 ± 0.41	95.63 ± 0.52
	b	96.74 ± 0.48	97.53 ± 0.29	95.80 ± 0.55	97.66 ± 0.32
0	a	97.02 ± 0.37	97.26 ± 0.20	96.82 ± 0.20	97.32 ± 0.22
	b	97.47 ± 0.21	97.91 ± 0.24	97.66 ± 0.28	97.63 ± 0.20
5	a	97.86 ± 0.23	97.54 ± 0.29	97.54 ± 0.05	97.63 ± 0.20
	b	98.10 ± 0.24	98.02 ± 0.20	98.11 ± 0.20	98.11 ± 0.19
10	a	97.86 ± 0.23	97.59 ± 0.19	97.66 ± 0.21	97.71 ± 0.19
	b	98.18 ± 0.21	98.22 ± 0.19	98.20 ± 0.18	98.24 ± 0.19

These show the $F1$ scores ($\%$) of HSS at different noise conditions. The notation 'a' feature represents the homomorphic, Hilbert, PSD, and wavelet envelopes used in David's work [5]. The 'b' features are the proposed HEoDF and Hilbert envelope after the dual filtering process, modified PSD, and wavelet envelopes.

feature, the proposal is compared with the existing features [5]. Their outcomes are shown in Table 5.2.

The performance scores shown in Table 5.1 and Table 5.2 illustrate that incorporating the multi-modal distribution model in HSMM has improved the segmentation accuracy especially in heart sound signals with considerably large heart rate variation. The incorporation of the dual-filter before feature extraction has also improved the segmentation of noisy recordings. This achievement has been elaborately discussed as under.

5.4.1 Effect of multi-centroid duration model

An example of HSS mentioned in [5, 16] is shown in Fig. 5.4. The corresponding duration model $p_i(d)$ for each state i is depicted in Fig. 5.5. In this existing method, the diastolic duration distribution is modeled such that the expected state duration is within the model permissible limit defined by $\mu d_i \pm \sigma d_i$. This duration distribution $p_i(d)$ is non-zero within this interval and zero elsewhere. Taking the value of σd_{siDia} determined by Eq. (5.4), the model distribution will ignore any abnormal state duration outside the stipulated interval. Therefore, the state duration will be forcibly selected within the model limit by the maximum likelihood criterion. If a subject suffers from abnormally large heart rate variation, the model may incorrectly estimate diastole interval.

In the proposed model, the silent diastolic duration d_{siDia} is estimated from every instance covering all potential duration values. Also, defining the model boundaries between $\text{dmin}_{\text{siDia}}$ and $\text{dmax}_{\text{siDia}}$ operates the estimation process within this adequate limit. Each weighted mode of the multi-modal distribution can be considered as a search space where the actual duration is expected. The true d_{siDia} is close to any of the modes represented by the distribution peaks. The final value is derived cumulatively from the probability of past state sequence, the transition probability a_{ij}, duration probability, and the probability of the present observation sequence, as defined by Eq. (2.54). This step ensures the derived state duration automatically adjusts itself to its best value. The modified duration model is illustrated in Fig. 5.7 and the outcome of the segmentation is shown in Fig. 5.6.

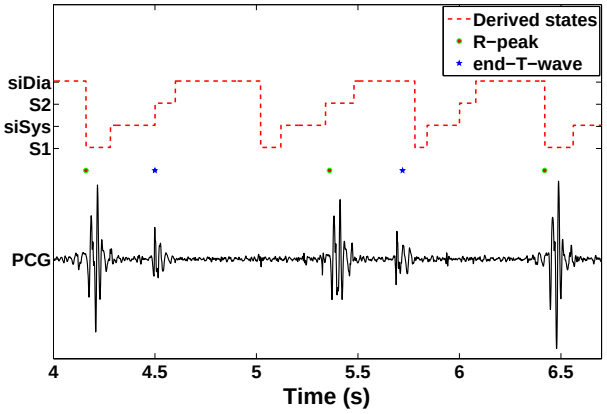

Figure 5.4: The derived state labels of a PCG using the existing LR-HSMM algorithm.

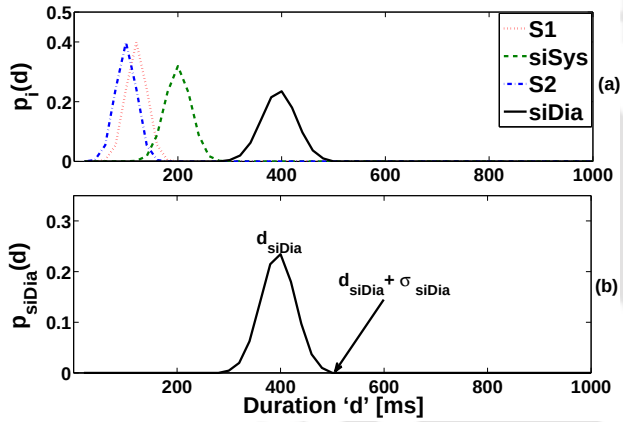

Figure 5.5: (a) The duration density $p_i(d)$ for each state i. (b) The density $p_{\text{siDia}}(d)$ using only mean value μd_{siDia}.

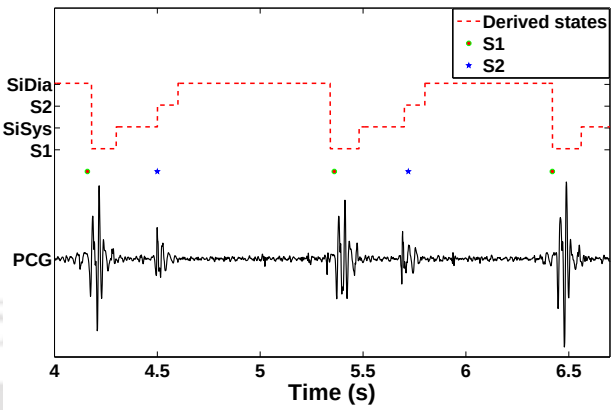

Figure 5.6: The derived state labels of a PCG using proposed duration model.

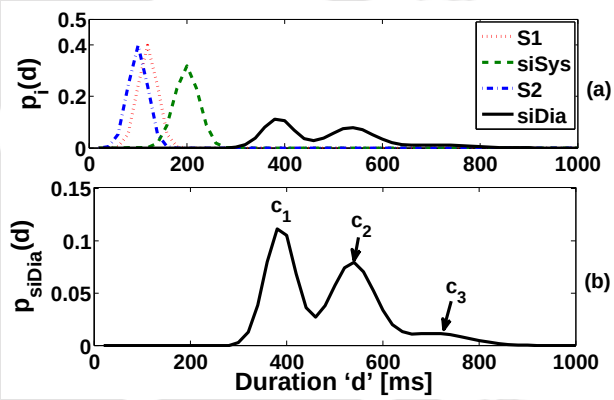

Figure 5.7: (a) The proposed duration density $p_i(d)$ modeled for each state. (b) The density $p_{\text{siDia}}(d)$ is distributed across the centroid locations c_1, c_2 and c_3.

5.4.2 Effect of TVF denoising

The dual filtering process of the standard band-pass filter (BPF) and TVF filter has considerably suppressed the noise elements that may be present in the silent interval between the S1 and S2 sounds in PCG. Analysis of the noisy PCG has shown its superior denoising capability compared to BPF, as shown in Fig. 5.8.

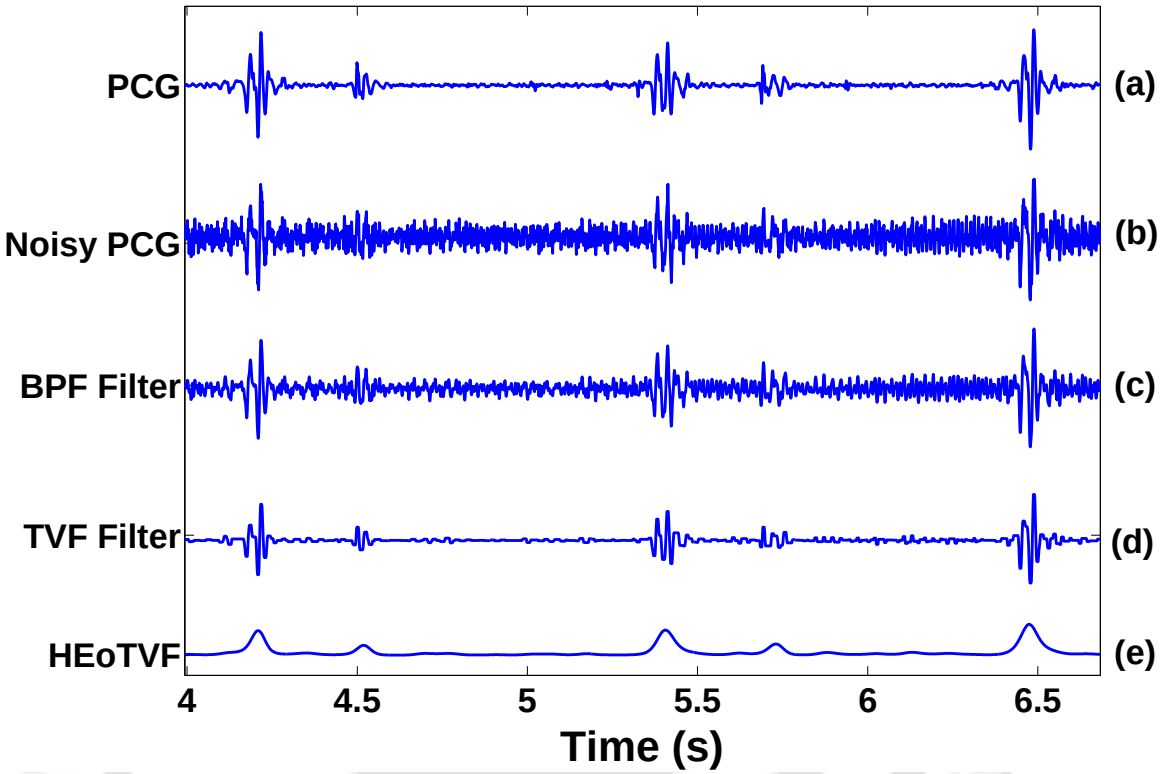

Figure 5.8: Example of HEoDF envelope for noisy PCG (SNR -5 dB) after dual filtering process.

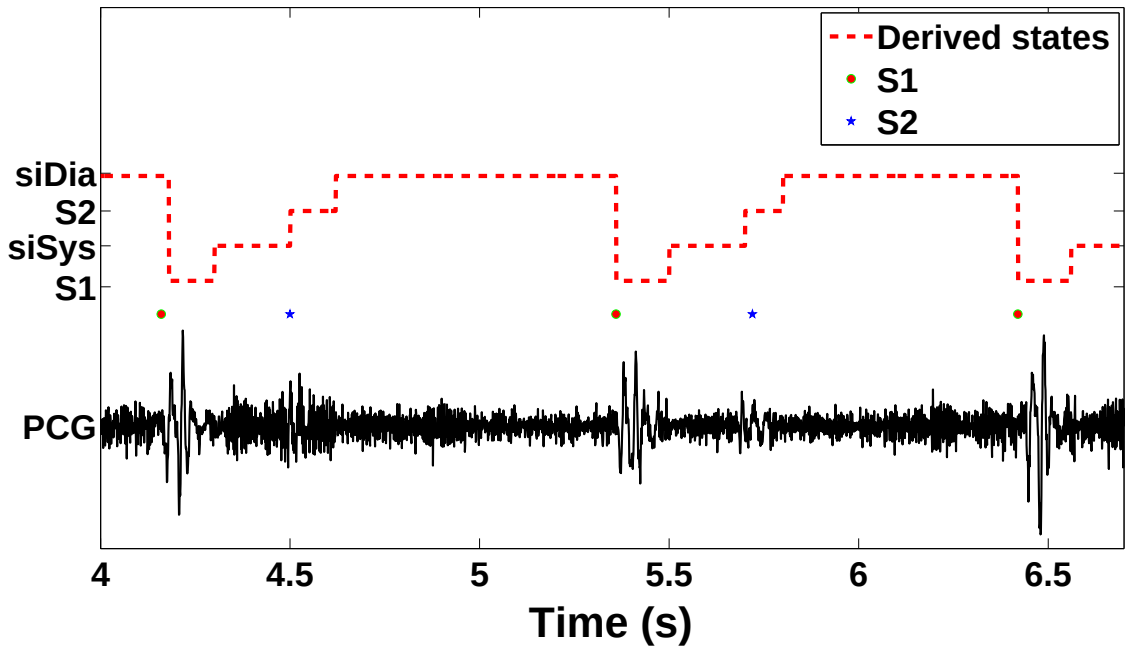

Figure 5.9: A segmented noisy PCG using the proposed algorithm.

This is because TVF is a piecewise constant smoothing technique. It smooths out low amplitude highly varying signal components and preserves the discontinuities in the signal. When used after processing with BPF, it suppresses leakage frequency components. Therefore, the HEoDF envelope extracted from the resulting filtered signal will hold more accurate signal characteristics than the normal homomorphic envelope, giving us distinct peaks and uniformly attenuated envelope segments that improve the discrimination of FHS from the silent intervals. Also, the duration parameters estimated from this envelope will be precise. The combination of the HEoDF feature with the proposed duration model has yielded considerable segmentation accuracy of 98.28 ± 0.19 Se and 98.45 ± 0.16 P_+ in Table 5.1. An example of a noisy PCG segmentation is shown in Fig. 5.9. The significance of this dual filtering technique is further illustrated by the performance scores at different degrees of additive noise in Table 5.2.

5.4.3 Effect of short duration test data

During recordings, the initial PCG is prone to the motion artifacts or noise generated during the stabilization of the recording instrument. Such noise usually has a high amplitude and may falsely appear in the HEoDF envelope even after the dual-filtering process. The outcome may cause an erroneous estimation of HCD. If the test data is too short, there is not enough HCD information to identify such errors, leading to an incorrect duration model. Usually, this issue is resolved by neglecting the first two heart cycles. Therefore, it is preferable to use longer PCG recordings ($>$ 5 HCD) for analysis.

5.4.4 Comparison with DRNN-method

Table 5.3 shows the comparison of results achieved with our proposed method and the DRNN-method. In contexts to the task (segmentation and identification of the four major FHS components) addressed in this work, our extended multi-mode duration model-based HSMM model exhibits much improvement over the existing one. If the problem statement is extended to identify the extra sounds such as S3 and S4 sounds, pathological murmurs and noise,

Table 5.3: Comparison of our proposed LR-HSMM extension with existing LR-HSMM [5] and DRNN-method [6].

Features	Classifier	Se	P_+	F_1
HEoDF, modified PSD, Hilbert and Wavelet	Proposed	98.28	98.45	98.36
Hilbert, Homomorphic, PSD, Wavelet	LR-HSMM [5]	95.02	95.78	95.40
LMS, MFCC, Δ, Δ^2, Hilbert, Homomorphic, Hilbert and Wavelet	DRNN [6]	95.10	96.10	95.60

the HSMM based methods may not be feasible because this method predicts the state by maximization of likelihood scores of the expected events. Faulty prediction is always expected if the trained model of any particular extra sound does not exist in the test PCG signal.

5.5 Summary

This chapter proposes a multi-modal diastolic duration model for the hidden semi-Markov model (HSMM)-based heart sound segmentation. The existing state duration model of HSMM is a single-mode distribution over a small range of variance and has ignored the possibilities of abnormal heart rate variation. The proposed model is developed by estimating all possible durations of the diastolic interval and clubbing them into smaller clusters by hierarchical agglomerative clustering based on Ward's method. Then, the modes are defined as the weighted Gaussian distribution of the centroids and the cluster variances. The multiple peak distribution yields a sharper gradient of likelihood around the expected modes and improves the discriminability of similar observations. Each mode acts as a reference point for the HSMM to determine the origin of the hidden state and adjust the corresponding state duration based on the maximum likelihood criterion. The proposed modification improves the segmentation accuracy yielding an average F_1 score of 98.36 ± 0.43. Incorporating the

proposed dual filtering process that combined the band-pass filter with the total variation filter has also improved the system's overall performance in dealing with noisy phonocardiograms.

6

Classification of heart sounds

Contents

Chapter 3, Chapter 4, and Chapter 5 discuss preprocessing, envelope enhancement, and heart sound segmentation tasks. A few novel algorithms were evaluated and found to be satisfactory. In this Chapter, some popular frequency domain features, including the Mel frequency cepstral coefficients (MFCC), rhythm patterns (RP), sub-band energies (SBE), and inter-segment correlation coefficients, have been explored as potential diagnostic features for the classification of PCG signal. As a preliminary study, these features are evaluated to classify heart sound into broad categories, namely normal, murmur, and noisy.

After going through the earlier chapters, it is clear that the signal characteristic of normal heart sound is relatively straightforward. The S1 and S2 sounds are short burst sounds that are temporally separated by silent intervals denoting the systole and diastole phases of a cardiac cycle. The consented frequency band of these fundamental heart sound signals ranges between 20 Hz to 200 Hz, with center frequency at 50 Hz [5, 18–20]. In pathological conditions, abnormal sounds may be introduced to the heart sound. In the case of valvular heart disease and septal defects, murmurs are produced in the heart sound due to the turbulent flow of blood through a narrow opening under high pressure. Experts diagnose these cardiac problems by listening to the heart sound via the auscultation process [67, 115]. The finding is used as an initial clue for further diagnosis. The standard technique to confirm the valvular and septal dysfunction is by echocardiography. Alternate means of diagnosis are electrocardiogram (ECG), ultrasound, magnetic resonance imaging (MRI), computerized tomography (CT) scans, angiography, etc., which are more expensive approaches.

In an automated diagnostic system, the heart sound signals are analyzed for sound pitch, duration, temporal location, sound intensity configurations, and spectro-temporal characteristics of heart sound and the abnormal sounds [100]. Some clinical traditions that are formulated into developing diagnostic systems are (i) sequential temporal patter on S1 and S2 sounds, (ii) localized nature of innocent murmurs at early to mid-systolic, and (iii) intensity configuration (crescendo, decrescendo, or crescendo-decrescendo) of murmur resembling the pressure built up in the heart chamber. The temporal pattern is realized in many works by analyzing envelope features. In classification processes, temporal sequence tagging is

possible using the HSMM model or LSTM networks [6, 66, 68–70]. In some works, statistical measures such as skewness, kurtosis [39, 116], and complexity measures [58, 59, 117] such as sample entropy are used to quantify signal layouts. Researchers around the globe are exploring various feature extraction methods and machine learning platforms to get a robust classification of the heart sound. The complex nature of the heart sound signal and variability of the signal characteristics for different subjects pose a difficult challenge to solve the classification problem. The PhysioNet's Computing in Cardiology Challenge 2016 widely discussed the potential aspects of different assessment and classification methods to detect abnormalities accurately in clinical applications. FHS is the most discernible sound in a cardiac cycle following a generic indubitable pattern. This enables researchers to design and compute many heart sound signal envelopes as features. But the envelope feature is susceptible to noise interferences. By applying a suitable filter system, the principle attributes of the envelope can be improved to some extent.

The spectral properties of the heart sound signal have been well established and used for the classification of generic classes representing FHS components [18]. After extensive analysis of PCG data from various research groups recorded from subjects with different age groups and health conditions, it is observed that the frequency band of the heart sound components are narrow and may overlap with murmurs, physiological artifacts, and non-physiological events. The typical frequency ranges for each heart sound component are discussed in Chapter 1, Table 1.1. The S1 and S2 sounds have relatively identical morphology and a similar frequency range (20-200 Hz). The gallop S3 and the S4 sounds manifest identically with a frequency range of 15-65 Hz which overlap with FHS. They are differentiated by determining the temporal allocation of the sounds in the cardiac cycle. On the other hand, murmurs exhibit diverse frequency ranges which can be as high as 700 Hz depending on the pathology. This non-definitive frequency composition poses difficulties to draw a standard skeleton of frequency elements that will separate heart sound components from each other, and from pathological sounds or artifacts. In some cases, the morphology of some noise may resemble the heart sound and pathological sound that makes the classification process even

more challenging.

The classification algorithm by Potes *et al* [118] claimed to have the best $\mathrm{M\_Acc}$ of 86.02% in identifying normal and abnormal heart sounds. The $\mathrm{M\_Acc}$ is calculated as a mean of sensitivity (Se) and specificity (Sp). The features used in their work include the mean and standard deviation of the following parameters.

i PCG intervals related: RR intervals, S1 intervals, S2 intervals, systolic intervals, diastolic intervals, the ratio of the systolic interval to RR interval of each heartbeat, the ratio of the diastolic interval to RR interval of each heartbeat, the ratio of the systolic to the diastolic interval of each heartbeat;

ii PCG amplitudes related: ratio of the mean absolute amplitude during systole to that during the S1 period in each heartbeat, the ratio of the mean absolute amplitude during diastole to that during the S2 period in each heartbeat, skewness of the amplitude during S1 period in each heartbeat, skewness of the amplitude during S2 period in each heartbeat, skewness of the amplitude during systole period in each heartbeat, skewness of the amplitude during diastole period in each heartbeat, kurtosis of the amplitude during S1 period in each heartbeat, kurtosis of the amplitude during S2 period in each heartbeat, kurtosis of the amplitude during systole period in each heartbeat, kurtosis of the amplitude during diastole period in each heartbeat.

iii Frequency domain feature: the median power across nine frequency bands (i.e. 25-45, 45-65, 65-85, 85-105, 105-125, 125-150, 150-200, 200-300, 300-400 Hz) corresponding to the S1, S2, systole, and diastole states of each cardiac cycle. In addition mean of 1-13 mel-frequency cepstral coefficients (MFCC) extracted from each state and each cardiac cycle were also included [118].

Zabihi *et al* [119] have used the ensemble of SVM and acquired an $\mathrm{M\_Acc}$ of 85.90%. The features used are 10^{th} order linear prediction coefficients (LPC), entropy, 14 MFCC coefficients, details and approximation coefficients of 5 level wavelet decomposition, and power spectrum of the signal.

In [6], E. Messner *et al* [6] have used a deep recurrent neural network as an alternate heart sound segmentation algorithm. They have shown the versatility of the classifier to detect not only FHS but additional sounds such as galloping S3 and S4 and some murmurs. The features used are 41-bin log-magnitude spectrogram, 12 MFCC coefficients, and its derivatives with the envelope features used in [5]. They have achieved a comparable F1 score of 95.6% with the state of the art algorithm.

The existing heart sound classification methods mostly focus on automated identification of abnormal heart sound from normal sound. However, the classification of abnormal heart sound is a broad category consisting of different pathological origins and having complex temporal and frequency characteristics. Computing feature sets to accommodate all the abnormalities is a difficult task. For example, the feature set to determine abnormal cardiac rhythm may not be feasible to detect gallop sounds or abnormal murmurs. Therefore it will be better to divide the classification problem into a much smaller problem. In this Chapter, we try to evaluate the normal heart sound, abnormal murmurs and noise using the sub-band feature and MFCC features.

Among the sub-band features and their derivatives, the MFCC feature is frequently used as a diagnostic feature to classify heart sound signals [6, 100, 120]. But, there is no consistency in the number of coefficients or frequency banks in the reported literature dedicated to the classification of heart sound signals. Therefore, comprehensive studies are required to examine the responses of changing the number of coefficients and filterbanks for detecting pathological murmurs. We have also introduced the sub-band energies (SBE) feature calculated across 25-50 Hz, 50-100 Hz, 100-250 Hz, 250-500 Hz, and 500-1000 Hz. Each sub-band energy implicates the energy of frequency components that may occur dominantly in a certain type of heart sound. Using larger sub-band intervals at higher frequency bands helps in boosting the cumulative information of low-intensity high-frequency signal components. The performance of these features is compared with the existing SBE features introduced by [118] and [16], which is evaluated using Support Vector Machine (SVM).

6.1 Features derived from heart sound recordings

The features used for this study are (a) the perceptual MFCC, (b) the acoustic rhythm pattern (RF), (c) the SBE feature, and (d) the intra-component correlation. The performance of these features is evaluated using SVM. They are briefly discussed further.

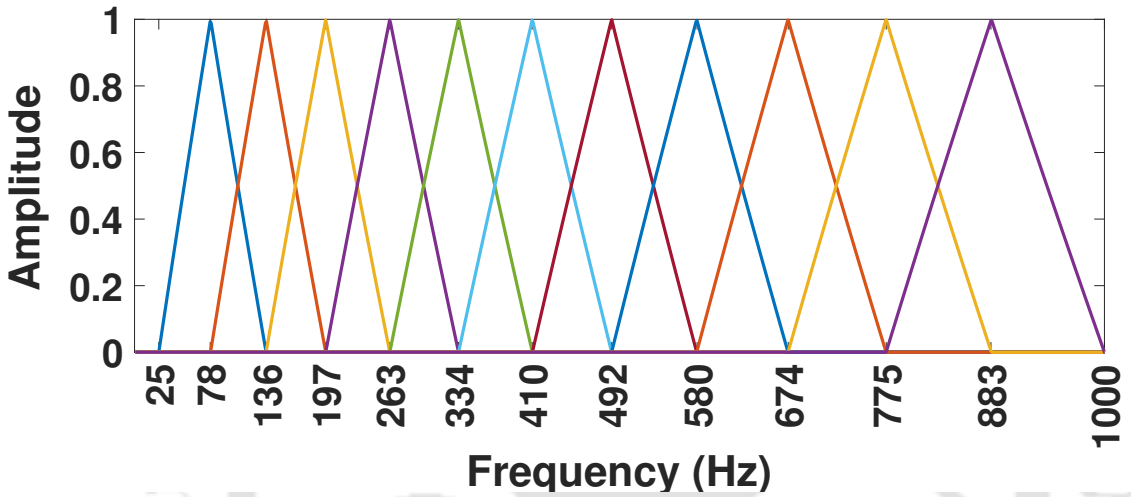

Figure 6.1: Mel filter bank.

Mel frequency cepstral coefficients (MFCC)

The MFCC feature has been used as an effective descriptor in various acoustic pattern recognition tasks including speech and speaker recognition [121–123]. The application of psychophysical pitch scales/Mel frequency warping is extended to classifications of the heart sound signal due to its perceptual competence. The MFCC coefficients are calculated by taking a discrete cosine transform of logarithmic power at each Mel frequency bank. The relation between the Mel frequency and the actual frequency of a signal is given by the Equation 6.1.

$$\mathrm{mel}(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right). \tag{6.1}$$

Considering the sampling frequency of the PCG dataset as 2 kHz, the number of filter-banks is taken as 11 so that it approximates the frequency-band description of 26 filter-banks

at 8 kHz sampling frequency used in speech processing applications, shown in Fig. 6.1. The filterbank starts at 25 Hz and ends at 1000 Hz.

Rhythm Patterns (RP)

The RP was extensively used as an acoustic feature for the classification of music. It captures the specific loudness sensation using Bark scales [124]. The extraction of RP involves the translation of cumulative spectrum energy across each Bark scale critical band into Phon/Decibel scale (E_{dB}). The process generates a sound intensity curve. Then the specific loudness sensation of a signal perceived by the human ear, known as the Bark scale sonogram (E_{Sone}), is computed using the Equation 6.2 below.

$$E_{\mathrm{Sone}} = 2^{\frac{E_{\mathrm{dB}}-40}{10}}. \tag{6.2}$$

The process transforms the Phon scale to the Sone scale. The final RP feature is obtained after Gaussian smoothing of the E_{Sone}. The Bark scale critical bands [7] are illustrated in Table 6.1. However, the measure of loudness is subjective and differ from person to person. A sound generated at equal deciBel but different frequencies are perceived to have unequal intensity.

Table 6.1: Bark scale critical band [7].

Number	Center Freq.	Cut-off Freq.	Bandwidth
		20	
1	60	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150

Sub-band energy (SBE)

This feature is derived from the sub-bands feature used in [16] by Schmidt *it al*. They have calculated SBE as the energy between 25-50 Hz, 50-100 Hz, and 100-150 Hz. The spectral ranges are limited and confined to the frequency composition of the normal heart sound only. In our proposed SBE based method, the frequencies of sub-bands are extended to higher frequency ranges covering the spectral characteristics of pathological murmurs. It is tabulated below.

Table 6.2: Extended frequency ranges of sub bands.

Number	Center Freq.	Cut-off Freq.	Bandwidth	Definition
		25		
1	37.5	50	25	Lower frequency range of S1 and S2.
2	75	100	50	Higher frequency range of S1 and S2.
3	175	250	150	Lower frequency range of murmurs.
4	375	500	250	Higher frequency range of murmurs.
5	750	1000	500	Remaining frequency bands dominated by noise.

The energy of the log magnitude spectrogram (LogMS) at each sub-band is taken as the final SBE feature. The advantage of the proposed SBE is that each frequency band segregates the frequencies of different heart sound components, abnormalities and noise in a PCG signal and help in evaluating the dominant signal and classify the heart sound recording. Also, providing larger frequency intervals improve the quantification of low amplitude high-frequency components of murmurs and noise signals.

Intra-segment correlation

For a healthy heart sound signal, spectral analysis shows a close resemblance between S1 and S2 sounds. The spectrum of the remaining systole and diastole intervals usually are sparsely distributed with relatively weak intensity. The frequency composition in these intervals depends on the noise and artefacts, introduced during recordings either from the ambient or from the recording device. Therefore, the noise or the pathological murmurs

are clearly visible in the systole and diastole intervals. This affects the spectral intensity of the characteristics frequency components in the signal segment. To incorporate these relations, the correlation coefficients for pairs of (a) S1 and systole, (b) S1 and diastole, and (c) systole and diastole are calculated. Correlations of the S2 segment with others are omitted because the intensity of the S2 sound is relatively inconsistent compared to the S1 sound in the present dataset. In addition, the S1 and S2 sound have similar morphology and their derived correlation coefficients bear a redundant observation. Since in the time domain, the waveform of PCG does not give better correlations and for this reason, the frequency domain representation of the signal is used to calculate the intra-segment correlations. For each heart sound segment pair, the subband energy feature across frequency bin size, as in Table 6.2, are calculated to approximate the characterizable frequency distribution. From the result, the correlation coefficients are measured.

6.2 Evaluation process

The procedural protocol for the experiment and evaluation is shown in Fig. 6.2. It consists of two basic modules: training and testing. The denoising step is omitted as the classification process will try to detect the noise present in the PCG signal. However, the amplitudes of all the PCG signals are scaled using mean-max normalization before further processing.

The relevant features as discussed above are extracted using feature extraction block in both the training and testing phase. The SVM is trained using the SVM training model block as shown in Fig. 6.2. In the testing phase, the features derived from the PCG signal is fed to the SVM classifier to classify the signal into normal category or murmur category or a noisy PCG signal. Each of the modules is discussed further as follows.

6.2.1 Dataset used for evaluation

There is no dedicated database of clean normal PCG signals or signals with murmurs but without any noise interferences. The database available in the ‘classifying heart sounds challenge’ sponsored by PASCAL [74] provides PCG signals categorically. It consists of two

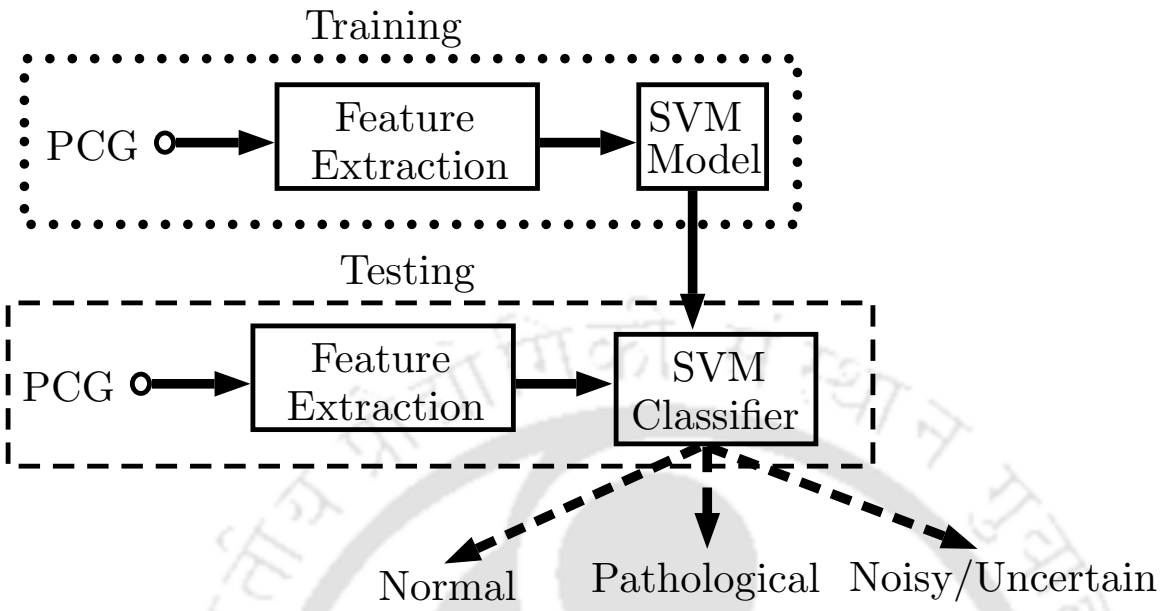

Figure 6.2: Block diagram of heart sound classification scheme.

datasets: (a) Dataset A and (b) Dataset B. Each of the datasets is sub-categorised as normal, murmurs, extra heart sound, and artifacts. The normal category is recorded from healthy subjects. Some of them may be corrupted with background noise (from traffic to radio), physiological sound and artifactes from recording devices (movement of microphone, and connecting wires). In the murmur category, the types of murmurs are not specified. However, it contains a wide range of murmurs having temporal locations in systole or diastole intervals and with varying degrees of intensities. Some of the recordings are noisy. The number of PCG signals in each dataset are tabulated in Table 6.3. For details please refer [74].

Table 6.3: Database for the classification of heart sounds by PASCAL.

	Dataset A.	Dataset B.
Normal	31	200
Murmur	34	66
Extra heart sound	19	NA
Extra systole	NA	46
Artifacts	40	NA

For the experimental analysis, the relatively noise-free datasets of normal and murmur classes are assembled from the above dataset A and B by choosing the PCG signal that

has cleaner systole and diastole intervals by inspecting the signal visually and through audio perception. The final dataset contains 62 files of normal category and 81 files of murmur category. The signals are then resampled to 2kHz. The noisy PCG data is created by introducing additive noise to the normal PCG signal at different noise levels. The description of additive noise is already been discussed in Chapter 2 Section 2.1.

The PCG signals are also labelled for their fundamental heart sound segments (S1, systole, S2 and diastole) by using the HSS algorithm which has been discussed in Chapter 5. The segmented labels are visually inspected, and corrections were made for any misalignment of the segment labels.

6.2.2 Feature extraction

For simplicity, the feature is extracted from the initial segment of the PCG signal with a length equal to one HCD (heart cycle duration). The test features are as follow:

- the MFCC feature is extracted by changing the number of coefficients from 3, 4, 5, 6 and 11 while keeping the number of filterbanks fixed at 11.
- the proposed SBE feature is extracted from both magnitude spectrogram and log-magnitude spectrogram.
- using the heart sound segmented labels, the four fundamental segments are identified. For each segment, the proposed SBE is estimated. Then the correlation coefficients for pairs: (a) S1 and systole, (b) S1 and diastole (c) and systole and diastole are calculated.

6.2.3 Support vector machine (SVM)

The SVM is a learning machine meant for binary classification. It works based on kernel function ($\phi(\mathbf{x}_i)$) that maps the data points into the higher dimension to obtain linear separability of the feature matrix. The separation of feature space corresponding to positive or negative class is achieved by maximization of the class margin and simultaneously minimizing the

training error ε_i. This is calculated as follows.

$$\min\left(\frac{1}{2}\mathbf{w}^2 + C\sum_{i=1}^{N}\varepsilon_i\right) \tag{6.3}$$

with constrains:

$$y_i\left(\mathbf{w}^T\ \phi(\mathbf{x}_i) + b\right) \geq 1 - \varepsilon_i \tag{6.4}$$

$$\varepsilon_i \geq 0, \qquad i = 1 : L \tag{6.5}$$

where 'w' is a direction vector, 'C' is a regularization parameter that provides a trade-off between the margin and the training error, $y_i \in (1, -1)$ defines the class label, and x_i is a feature vector. This is solved as a quadratic optimization problem based on Karush Kuhn Tucker (KKT) condition.

$$\min\left(Q(\Lambda) = \sum_{i=1}^{L}\alpha_i - \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}\alpha_i\alpha_j y_i y_j \phi(\mathsf{x}_i, \mathsf{x}_j)\right) \tag{6.6}$$

constrains:

$$0 \leq \Lambda \leq C \tag{6.7}$$

$$\Lambda^T Y = 0 \tag{6.8}$$

where $\Lambda = [\alpha_1, \alpha_2, ..., \alpha_L]$ is the non-negative Lagrange multiplier. In this work the radial basis kernel function (RBF) is implemented, defined in Equation 6.9. For further details, we may refer [125, 126].

$$K(\mathsf{x}_i, \mathsf{x}_j) = \exp\left(-\frac{\|\mathsf{x}_i - \mathsf{x}_j\|^2}{2\sigma^2}\right) \tag{6.9}$$

In order to use SVM for multi-class classification, the problem is simplified by splitting the task into a series of binary problems. The method incorporated is the one-versus-rest (1VR) classification. A binary classifier is trained on each class against the remaining classes. Then the predictions are made based on the model that has the most confidence level.

6.3 Results and discussion

In the evaluation process, the leave-one-out cross-validation (LOOCV) is used to estimate the performance of the presented classification framework. In the LOOCV scheme, all observations (PCG signals) except one are used for training, and the one data left out is used for testing. This process is repeated for the total number of observations in the dataset. Though it is a computationally expensive procedure, it results in a reliable and unbiased estimate of model performance. It is fitting for evaluating a small dataset. The performance of the classification is measured as $\mathrm{M\_Acc}$, which is the mean value of sensitivity (Se) and specificity (Sp) [107]. The performance matrix is briefly explained in Chapter 2 Section 2.5. The findings are discussed in the following sub-sections as follows:

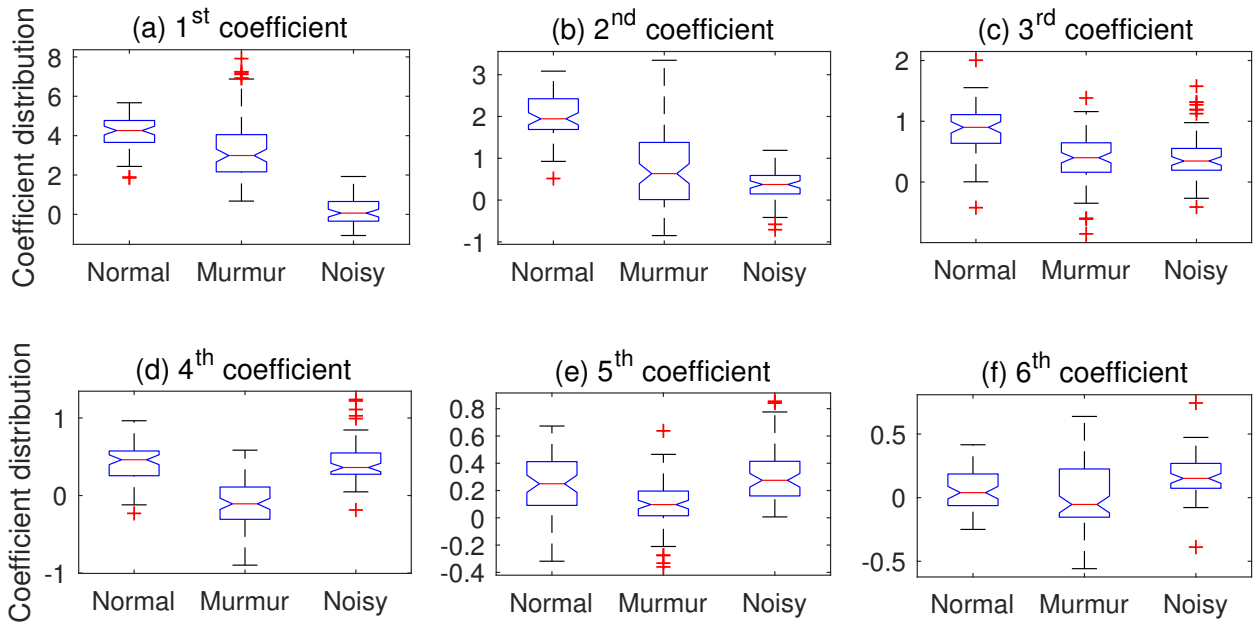

Figure 6.3: First six mfcc coefficients.

6.3.1 Performance of MFCC feature

The boxplots of the first six MFCC coefficients are illustrated in Fig. 6.3. Theoretically, a few lower-order MFCC coefficients represent the envelope of the spectral energy distribution.

This envelope may be correlated with the mechanical structure of the heart that produces the sound. Therefore, the first few coefficients have better characterization properties. The higher dimension of coefficients represents the spectral details. Including these coefficients may not improve the signal characterization. This may be because the model complexity increases. Or the number of data used for the experiment is too small to estimate the model parameters accurately.

Table 6.4 shows the performance scores of MFCC feature for a different number of MFCC coefficients. For 1-3 order MFCC coefficients, the $\mathrm{M\_Acc}$ values for normal, murmur and noise are 89.38%, 86.63% and 93.40%, respectively. As expected, the performance falls if higher-dimension MFCC coefficients (1-11) are taken. For the data set used in this study, taking 1-4 order MFCC coefficients is optimal and yields better performance with $\mathrm{M\_Acc}$ values for normal, murmur, and noise are 91.67%, 92.28% and 99.21%, respectively.

Table 6.4: Evaluation of the MFCC features at varying number of coefficients using SVM classifier. $\mathrm{M\_Acc}$ shown is in $\%$.

Types	Number of MFCC coefficients				
	1-3	**1-4**	**1-5**	**1-6**	**1-11**
Normal	89.38	91.67	90.53	89.61	81.05
Murmur	86.63	92.28	91.27	90.39	81.70
Noise	93.40	99.21	98.86	96.03	84.13

6.3.2 Performance of SBE feature

The proposed SBE feature is calculated from LogMS ($\mathrm{p\_SBE_{log}}$) and from the normal magnitude spectrum ($\mathrm{p\_SBE_{lóg}}$). After observing the performance of the features for the classification task, it is clear that introducing logarithmic scale in $\mathrm{p\_SBE_{log}}$ has improved the gross classification accuracy significantly. The signal energy distribution at different subbands is plotted in Fig. 6.4. In this feature, most normal heart sound signal energy is cumulated in the first and

second SBE coefficients. The SBE coefficients illustrated in Fig. 6.4 (a) and (b) suggest that in most cases, the principle frequency composition of the S1 and S2 signals are not affected by pathology and noise interference. The third and fourth SBE coefficient represents the energy of frequency components typical to pathological murmurs. Their coeficients distributions against the heart sound categories are shown in Fig. 6.4 (c) and (d). The ambient/acoustic noise usually has a relatively high-frequency content, and they are quantified in the fifth SBE value as shown in Fig. 6.4 (e).

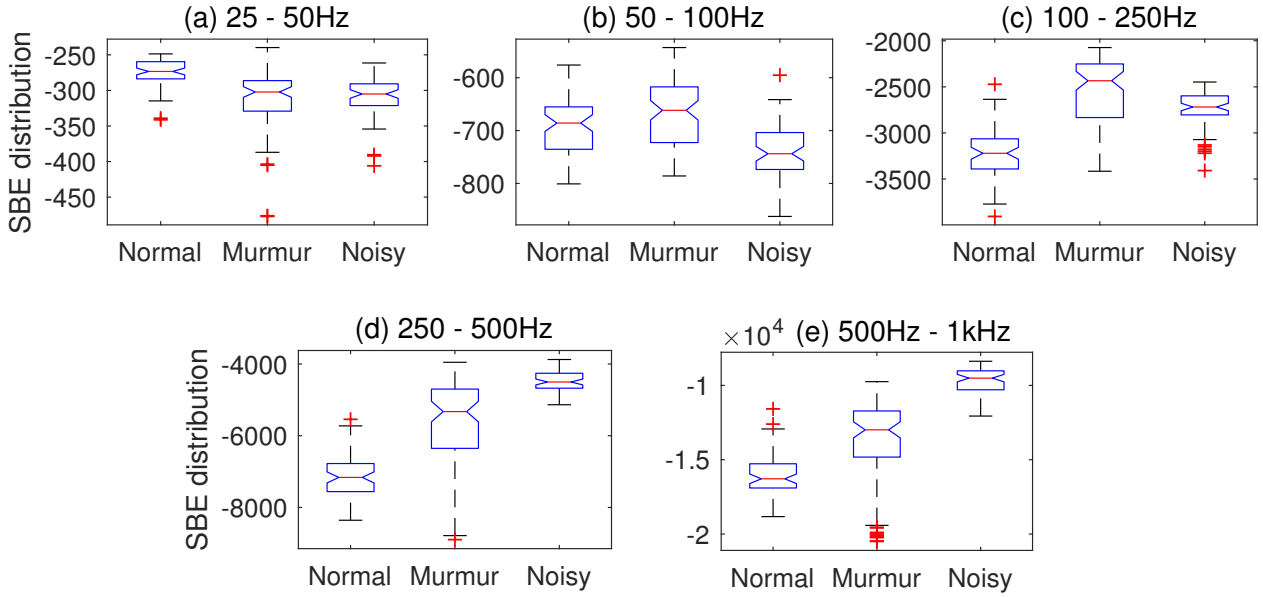

Figure 6.4: Sub-band energy feature defined over frequency ranges: (a) 25-50Hz, (b) 50-100Hz, (c) 100-250Hz, (d) 250-500Hz, and (e) 500-1000Hz.

The comparison of the proposed SBE with other existing features is illustrated in Table 6.5. The gross performance of the 1-4 order MFCC feature is better than other existing features. This indicates that the logarithmic scaling of frequency in Mel-filter banks and the mapping of energy distribution from log-magnitude spectrogram (LogMS) helps enhance the low amplitude high-frequency details of pathological sounds and noise. Using the proposed feature ($\mathrm{p\_SBE_{log}}$) for classification of normal, murmur, and noisy classes, the performance scores obtained are respectively 93.95%, 94.31% and 98.86%. After analyzing the evaluation results, it is observed that most features are good for classifying noisy PCG

signals. This is because the additive noise used for generating the noisy signal is mostly a high-frequency ambient noise. Also, the cause of poor classification accuracy for normal and pathological categories is because the frequency spectrum of the normal heart sound (NHS) and pathological murmurs are close or overlapping one another, rendering it difficult to segregate.

Table 6.5: Comparison of the proposed SBE with 1-4 MFCC and existing features using SVM classifier. $\mathrm{M\_Acc}$ value shown is in $\%$.

Types	Features							
	1-11 MFCC	**1-4 MFCC**	**SBE [16]**	**SBE [118]**	**41 LogMS [6]**	**RP [124]**	$\mathrm{p\_SBE_{log}}$	$\mathrm{p\_SBE_{log}}$
Normal	81.05	91.67	71.68	84.33	58.13	91.52	86.95	93.95
Murmur	81.70	92.28	77.07	86.68	55.47	89.15	84.35	94.31
Noise	84.13	99.21	84.62	86.51	56.35	97.82	95.68	98.86

The proposed SBE as the diagnostic feature is limited only to classify the normal (clean), pathological (murmurs), or noisy categories of PCG signals. Because taking a large frequency band to calculate SBE is not capable of providing detailed spectral characteristics of the signal. Instead, it approximates the cumulative information of the signal within the sub-band. The proposed features may perform better than the conventional features for classifying heart sound into broad classes. But it lags details for classifying murmurs and identifying their pathological sources. There is still a need for a more rigorous study of characteristic frequency properties of pathological murmurs to develop feasible subband features.

6.3.3 Correlation coefficients

The intention of measuring the systole and diastole correlation coefficients is to detect the dissimilarities in the two segments. For normal and noisy categories, it is expected that the two segments will have similar signal nature and yield a high coefficient value. The two segments will never be identical in the murmur category and always result in a smaller value. But some of the coefficient values may show a high correlation if the murmur sounds have

low intensity, such as benign murmur. The distributions of correlation coefficients are plotted in Fig. 6.5. The box plots in Fig. 6.5 (a) and (b) show a clear distinction between normal and noisy categories. In the case of the murmur category, murmur sound may appear in either systole or diastole intervals. As a result, the correlation coefficients calculated from pairs S1 versus systole (S1VsSys) and S1 versus diastole (S1VsDia) will significantly vary. Individually, they do not make any sense. On observing both the coefficient values simultaneously, it is realizable that the PCG signal with murmur will produce either a large S1VsSys coefficient and a small S1VsDia coefficient value or vice versa depending on the location of murmur sound. Therefore, these correlation coefficients may be used to characterize the murmur sound.

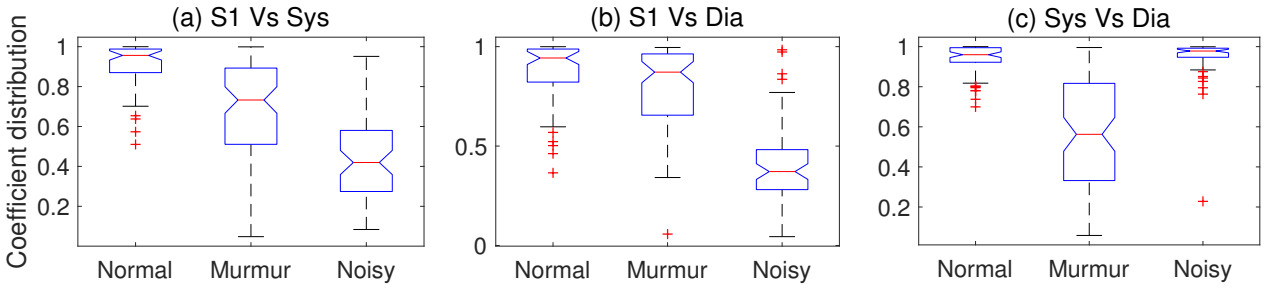

Figure 6.5: Correlation coefficients measured against the pair of (a) S1 and systole, (b) S1 and diastole, and (c) systole and diastole.

6.3.4 Performance with combination of features

The classification results of different combinations of feature sets are illustrated in Table 6.6.

Combining the 1-4 MFCC with the correlation coefficient of systole and diastole segments (CCSysVsDia) yields better performance over other feature combinations. The $\mathrm{M\_Acc}$ values for these features are 96.23%, 95.94% and 99.21% for normal, murmur, and noisy heart sounds, respective. Introducing CCSysVsDia with the proposed SBE does not significantly differ the classification performance over the combination with 1-4 MFCC feature. For this combination, the $\mathrm{M\_Acc}$ values are found 96.92%, 96.78% and 98.41% for normal, murmur, and noisy categories, respectively. But, the overall performance of all features combination is better than the best M Acc of 86.02% claimed by Potes *et al.* [118] in the PhysioNet/Computing in

Table 6.6: Performance of classification using combination of different features set. M_Acc values in %.

Feature combinations	Normal	Murmur	Noisy
SBE + 1-4 MFCC	91.47	93.96	98.86
all CC + 1-4 MFCC	95.09	91.53	93.30
SysVsDiaCC + 1-4 MFCC	96.23	95.94	99.21
all CC + SBE	95.23	95.59	96.03
CCSysVsDia + SBE	96.92	96.78	98.41
CCSysVsDia + 1-4 MFCC + SBE	91.82	91.40	93.65

all CC = Correlation Coefficients of pairs: S1 vs systole, S1 vs diastole, and systole vs diastole.
CCSysVsDia = Correlation Coefficient of pair systole vs diastole.

Cardiology Challenge 2016. Thus, it is found that the systole and diastole segment correlation coefficients with either the proposed SBE feature or the 1-4 order MFCC feature give better performance accuracy. By comparison of computational expenses to generate the feature, the proposed SBE feature is better.

6.4 Summary

In this Chapter, we have explored the MFCC, RP, SBE, and intra-segment correlation coefficients for possible classification of heart sound signals into broad categories of normal, murmur, or noisy classes. We also proposed a new SBE feature whose frequency band represents the dominant frequency compositions of different heart sound components. The SVM classifier is used to evaluate the classification performance of the features. In the experiment, the individual features and the combination of features are used as an input to check which feature/feature set discriminates the heart sound better. The results show that the 1-4 order MFCC coefficients yield good classification accuracy. The RP feature also delivers comparable performance with the MFCC features. However, the proposed SBE shows slight improvement over the 1-4 order MFCC coefficients and is much better than the others. When comparing different feature combinations for classification performance, it is observed that the combination of correlation coefficient due to systole and diastole segment

and the proposed SBE feature or the 1-4 order MFCC feature is better than the rest of the other combinations.

This study has shown that the sub-band features and derivatives can produce good discrimination for heart sound classification. With the more rigorous investigation on the characteristics frequency distribution and determining the effective energy configurations of pathological heart sounds and murmurs, we can find an optimal sub-band and improve the quality of feature discriminability. That way, the SBE feature may be enriched and extended for pathology sound identification. This investigation is still in the initial stage. The analysis is restricted to the frequency domain features only. There is still room for improvement by exploring different domains to find discriminative diagnostic features.

7

Conclusions

Contents

In this thesis work, the heart sound recordings are denoised, signal envelopes are enhanced, and the FHS's are segmented. Finally, heart sounds are classified into broad categories: normal, murmur or noisy. In the preprocessing module, a new PCG denoising and modified HSS algorithm have been introduced. The HSS algorithm will segment a PCG signal into its four fundamental segments: S1, systole, S2, and diastole. The proposed denoising process is a hybrid of BPF/WT and OGS-TVF with sample entropy as a smoothing parameter. The HSS algorithm makes used on the multi-mode duration model-based HSMM classifier to segment PCG signals. The envelope feature used for HSS is enhanced using the logistic function amplitude moderation method. The S-curve of the logistic function is modelled to uniformly enhanced the S1 and S2 signal envelopes and suppressed the noise signals. This work also attempts to develop new diagnostic features based on spectral properties to identify the PCG signal. The work in each chapter of this thesis is summarized as follows.

- In Chapter 1, the physiology of the heart and the mechanism of production of heart sound have been introduced. The chapter also discusses the clinical significance of different heart sounds and associated pathological ailments in the signal. In the last section of the chapter, the scope of the work is presented.

- In Chapter 2, a related review on the various modules in the segmentation and classification algorithm used in the thesis work is discussed. First, it discusses the advantages and disadvantages of the denoising methods based on the wavelet transform (WT) and the total variation filter (TVF), both standard and overlapping group sparsity approach. In the feature extraction module, a related review of the existing envelope extraction methods is included. Lastly, the potential of the HSMM classifier for the segmentation of heart sound signals is discussed. It also presents the difficulty of using a single-mode duration model for segmenting a PCG signal that has large heart rate variation. The motivation of this thesis is presented in the last section.

- In Chapter 3, the standard TVF and OGS-TVF are studied as a denoising method to suppress the noise in PCG signals. To adaptively penalize the noise in the signal, the

smoothing parameter of the TVF filter is derived from the complexity measure known as sample entropy. The proposed denoising algorithm combined BPF/WT and adaptive OGS-TVF so that the BPF/WT will suppress high-frequency noise, and if needed, the TVF will smooth the low-intensity residual noise. This preprocessing step is used to emphasize the S1 and S2 sounds to improve the HSS.

- In Chapter 4, the logistic function is used as an amplitude moderation method to uniformly enhance the S1 and S2 sounds. The asymptotic tail of the S-curve is calibrated to suppressed the noise signal in systole and diastole intervals. The scaling and shifting parameters are introduced into the function to adjust the S-curve with the nature of the PCG signal intensity distribution. The parameter values are derived from the individual signal using intensity histogram analysis.

  Using the proposed method, the S1 and S2 sound signals are emphasized yielding relatively uniform amplitudes compared to the systole and diastole intervals. The envelope derived from the transformed waveform improves the discrimination of the four fundamental cardiac intervals. The performance of HSS using the LFAM based envelope feature is better than the conventional envelope features.

- In Chapter 5, the multi-mode diastolic duration model is proposed for the HSMM based HSS algorithm. The durations of diastole intervals are estimated through autocorrelation analysis of the PCG envelope. It involves locating the time instance where the high peak of the autocorrelation function is observed. A simple hierarchical agglomerative clustering is used to group the estimated duration values to clusters such that the neighbouring cluster centroids are at least a systolic duration apart from each other. The process may end up with a single cluster or more than one. Based on the outcome, the final duration model is calculated by cascading the Gaussian distribution defined over each cluster. The proposed duration model greatly improves the segmentation accuracy of the PCG signal with large heart rate variation, and thus the overall performance of

the HSMM based HSS algorithm. The average F_1 score obtained is $98.36 \pm 0.43\%$.

- In Chapter 6, the proposed SBE feature calculated across frequency ranges dominantly associated with the healthy heart sound, murmurs and noise is evaluated for the classification of PCG signals. In addition, perceptual features such as MFCC, rhythm pattern (RP), other existing SBE features and the correlation coefficients of pairs: (i) S1 and systole, (ii) S1 and diastole, and (iii) systole and diastole are also tested. Comparing the results by taking individual feature for classification of PCG signals have shown that the proposed SBE feature and 1-4 order MFCC performed batter with average $\mathrm{M\_Acc}$ of 94.39% and 95.71% respectively compared to the remaining features. When the classification performance is tested for the combination of different feature sets, introducing the correlation coefficient of systole and diastole segments along with the proposed SBE feature gives a better performance with an average $\mathrm{M\_Acc}$ of 97.37%.

7.1 Scope for the future work

The following would be the directions for future work.

- The use of frequency-domain features shows potential for categorizing PCG signals into a broad category of healthy, murmur or noisy. However, the proposed features alone may fail to characterize certain murmurs, such as low-intensity benign murmur and murmurs whose frequency spectrum overlaps the normal heart sound. Introducing morphological features such as the signal intensity configurations and amplitude ratio of different segments may improve the classification performance.

- Bidirectional recurrent neural networks and HSMM classifiers have great potential to analyse the temporal relation of different heart sound components. Using the proposed features and other morphological features, the classifiers may be implemented to simultaneously segment and identify the murmurs and noise in the PCG signal.

- With rigorous analysis, the proposed diagnostic features can be extended for identifying the pathological origin of the murmurs.

References

[1] S.-W. Deng and J.-Q. Han, "Adaptive overlapping-group sparse denoising for heart sound signals," *Biomedical Signal Processing and Control*, vol. 40, pp. 49–57, 2018.

[2] H. Naseri and M. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Annals of biomedical engineering*, vol. 41, no. 2, pp. 279–292, 2013.

[3] M. Nabih-Ali, E.-S. A. El-Dahshan, and A. S. Yahia, "A review of intelligent systems for heart sound signal analysis," *Journal of Medical Engineering & Technology*, vol. 41, no. 7, pp. 553–563, 2017.

[4] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 609–612.

[5] 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, April 2016.

[6] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation—an event detection approach using deep recurrent neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964–1974, Sep. 2018.

[7] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Springer Science & Business Media, 2013, vol. 22.

[8] L. Bahekar, A. Misal, and G. Sinha, "Heart sound segmentation techniques: a survey," *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), e-ISSN*, pp. 2278–1676, 2014.

[9] S. Li, F. Li, S. Tang, and W. Xiong, "A review of computer-aided heart sound detection techniques," *BioMed Research International*, vol. 2020, 2020.

[10] K. Fox, J. S. Borer, A. J. Camm, N. Danchin, R. Ferrari, J. L. L. Sendon, P. G. Steg, J.-C. Tardif, L. Tavazzi, M. Tendera, *et al.*, "Resting heart rate in cardiovascular disease," *Journal of the American College of Cardiology*, vol. 50, no. 9, pp. 823–830, 2007.

[11] W. B. Kannel, C. Kannel, R. S. Paffenbarger Jr, and L. A. Cupples, "Heart rate and cardiovascular mortality: the framingham study," *American heart journal*, vol. 113, no. 6, pp. 1489–1494, 1987.

[12] J. E. Hall and M. E. Hall, *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.

[13] R. M. Rangayyan and R. J. Lehner, "Phonocardiogram signal analysis: a review." *Critical reviews in biomedical engineering*, vol. 15, no. 3, pp. 211–236, 1986.

[14] B. Karnath and W. Thornton, "Auscultation of the heart," *Hospital Physician*, vol. 38, no. 9, pp. 39–45, 2002.

[15] H. K. Walker, W. D. Hall, and J. W. Hurst, *Peripheral Blood Smear–Clinical Methods: The History, Physical, and Laboratory Examinations*. Butterworths, 1990.

[16] S. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiological measurement*, vol. 31, no. 4, p. 513, 2010.

[17] A. Waugh and A. Grant, *Ross & Wilson anatomy and physiology in health and illness*. Elsevier Health Sciences, 2014.

[18] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson, *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, no. 12, p. 2181, 2016.

[19] P. Arnott, G. Pfeiffer, and M. Tavel, "Spectral analysis of heart sounds: relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives," *Journal of biomedical engineering*, vol. 6, no. 2, pp. 121–128, 1984.

[20] M. Abella, J. Formolo, and D. G. Penney, "Comparison of the acoustic properties of six popular stethoscopes," *The Journal of the Acoustical Society of America*, vol. 91, no. 4, pp. 2224–2228, 1992.

[21] K. Maganti, V. H. Rigolin, M. E. Sarano, and R. O. Bonow, "Valvular heart disease: diagnosis and management," in *Mayo Clinic Proceedings*, vol. 85, no. 5. Elsevier, 2010, pp. 483–500.

[22] A. Freeman and S. A. LEVINE, "The clinical significance of the systolic murmur: A study of 1000 consecutive" non-cardiac" cases," *Annals of Internal Medicine*, vol. 6, no. 11, pp. 1371–1385, 1933.

[23] H. K. Walker, W. D. Hall, J. W. Hurst, and editors, *Clinical Methods: The History, Physical and Laboratory Examinations. 3rd edition.* Boston: Butterworths, 1990. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK201/

[24] A.-L. Noponen, S. Lukkarinen, A. Angerla, and R. Sepponen, "Phono-spectrographic analysis of heart murmur in children," *BMC pediatrics*, vol. 7, no. 1, p. 1, 2007.

[25] D. Balasubramaniam and D. Nedumaran, "Efficient computation of phonocardiographic signal analysis in digital signal processor based system," *International journal of computer theory and Engineering*, vol. 2, no. 4, p. 660, 2010.

[26] A. Sakai, L. P. Feigen, and A. A. Luisada, "Frequency distribution of the heart sounds in normal man," *Cardiovascular research*, vol. 5, no. 3, pp. 358–363, 1971.

[27] R. F. Rushmer, "Cardiovascular dynamics." *Academic Medicine*, vol. 36, no. 6, p. 742, 1961.

[28] R. J. Adolph, J. F. Stephens, and K. Tanaka, "The clinical value of frequency analysis of the first heart sound in myocardial infarction," *Circulation*, vol. 41, no. 6, pp. 1003–1014, 1970.

[29] E. Van Vollenhoven, A. Van Rotterdam, T. Dorenbos, and F. Schlesinger, "Frequency analysis of heart murmurs," *Medical and biological engineering*, vol. 7, no. 2, pp. 227–232, 1969.

[30] R. J. Lehner and R. M. Rangayyan, "A three-channel microcomputer system for segmentation and characterization of the phonocardiogram," *IEEE Transactions on Biomedical Engineering*, no. 6, pp. 485–489, 1987.

[31] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen, "Computer-based detection and analysis of heart sound and murmur," *Annals of biomedical engineering*, vol. 33, no. 7, pp. 937–942, 2005.

[32] C. Ahlstrom, P. Hult, P. Rask, J.-E. Karlsson, E. Nylander, U. Dahlström, and P. Ask, "Feature extraction for systolic heart murmur classification," *Annals of biomedical engineering*, vol. 34, no. 11, pp. 1666–1677, 2006.

[33] C. Ahlstrom, T. Länne, P. Ask, and A. Johansson, "A method for accurate localization of the first heart sound and possible applications," *Physiological Measurement*, vol. 29, no. 3, p. 417, 2008.

[34] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*. IEEE, 1997, pp. 105–108.

[35] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, "Neural network classification of homomorphic segmented heart sounds," *Applied Soft Computing*, vol. 7, no. 1, pp. 286–297, 2007.

[36] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, "A robust heart sounds segmentation module based on s-transform," *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 273–281, 2013.

[37] V. N. Varghees and K. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3861–3872, 2017.

[38] L. N. Sharma, "Teager-kaiser energy operator in hilbert space for heart sound segmentation," in *Biomedical Engineering (BME-HUST), International Conference on*. IEEE, 2016, pp. 8–13.

[39] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1138–1152, 2014.

[40] H. P. Sava and J. E. McDonnell, "Spectral composition of heart sounds before and after mechanical heart valve implantation using a modified forward-backward prony's method," *IEEE transactions on biomedical engineering*, vol. 43, no. 7, pp. 734–742, 1996.

[41] R. K. Sinha, Y. Aggarwal, and B. N. Das, "Backpropagation artificial neural network classifier to detect changes in heart sound due to mitral valve regurgitation," *Journal of Medical Systems*, vol. 31, no. 3, pp. 205–209, 2007.

[42] M. Akay, J. Semmlow, W. Welkowitz, M. Bauer, and J. Kostis, "Detection of coronary occlusions using autoregressive modeling of diastolic heart sounds," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 4, pp. 366–373, 1990.

[43] Y. M. Akay, M. Akay, W. Welkowitz, J. L. Semmlow, and J. B. Kostis, "Noninvasive acoustical detection of coronary artery disease: a comparative study of signal processing methods," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 6, pp. 571–578, 1993.

[44] J. Herold, R. Schroeder, F. Nasticzky, V. Baier, A. Mix, T. Huebner, and A. Voss, "Diagnosing aortic valve stenosis by correlation analysis of wavelet filtered heart sounds," *Medical and Biological Engineering and Computing*, vol. 43, no. 4, pp. 451–456, 2005.

[45] I. Gu, S. Kara, N. F. Gu, M. K. Kiymik, *et al.*, "Application of autoregressive and fast fourier transform spectral analysis to tricuspid and mitral valve stenosis," *Computer methods and programs in biomedicine*, vol. 49, no. 1, pp. 29–36, 1996.

[46] J. C. Wood and D. T. Barry, "Time-frequency analysis of the first heart sound," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 2, pp. 144–151, 1995.

[47] J. J. Lee, S. M. Lee, I. Y. Kim, H. K. Min, and S. H. Hong, "Comparison between short time fourier and wavelet transform for feature extraction of heart sound," in *TENCON 99. Proceedings of the IEEE Region 10 Conference*, vol. 2. IEEE, 1999, pp. 1547–1550.

[48] B. Ergen, Y. Tatar, and H. O. Gulcur, "Time–frequency analysis of phonocardiogram signals using wavelet transform: a comparative study," *Computer methods in biomechanics and biomedical engineering*, vol. 15, no. 4, pp. 371–381, 2012.

[49] D. L. Jones and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 40, no. 2, pp. 413–420, 1992.

[50] S. R. Messer, J. Agzarian, and D. Abbott, "Optimal wavelet denoising for phonocardiograms," *Microelectronics journal*, vol. 32, no. 12, pp. 931–941, 2001.

[51] S. Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1679–1687, 2008.

[52] A. A. Sepehri, A. Gharehbaghi, T. Dutoit, A. Kocharian, and A. Kiani, "A novel method for pediatric heart sound segmentation without using the ecg," *Computer methods and programs in biomedicine*, vol. 99, no. 1, pp. 43–48, 2010.

[53] S. Ari and G. Saha, "Classification of heart sounds using empirical mode decomposition based features," *International Journal of Medical Engineering and Informatics*, vol. 1, no. 1, pp. 91–108, 2008.

[54] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[55] J. Botha, C. Scheffer, W. Lubbe, and A. Doubell, "Autonomous auscultation of the human heart employing a precordial electro-phonocardiogram and ensemble empirical mode decomposition," *Australasian Physical & Engineering Sciences in Medicine*, vol. 33, no. 2, pp. 171–183, 2010.

[56] V. N. Varghees and K. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomedical Signal Processing and Control*, vol. 13, pp. 174–188, 2014.

[57] J. Pedrosa, A. Castro, and T. T. Vinhoza, "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 2294–2297.

[58] V. Nigam and R. Priemer, "Accessing heart dynamics to estimate durations of heart sounds," *Physiological measurement*, vol. 26, no. 6, p. 1005, 2005.

[59] J. Vepa, P. Tolay, and A. Jain, “Segmentation of heart sounds using simplicity features and timing information,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 469–472.

[60] L. Gamero and R. Watrous, “Detection of the first and second heart sound using probabilistic models,” in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 3. IEEE, 2003, pp. 2877–2880.

[61] A. D. Ricke, R. J. Povinelli, and M. T. Johnson, “Automatic segmentation of heart sound signals using hidden markov models,” in *Computers in Cardiology, 2005*. IEEE, 2005, pp. 953–956.

[62] D. Gill, N. Gavrieli, and N. Intrator, “Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model,” in *Computers in Cardiology, 2005*. IEEE, 2005, pp. 957–960.

[63] J. Oliveira, F. Renna, T. Mantadelis, and M. Coimbra, “Adaptive sojourn time hsmm for heart sound segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 642–649, March 2019.

[64] J. Oliveira, F. Renna, and M. Coimbra, “A subject-driven unsupervised hidden semi-markov model and gaussian mixture model for heart sound segmentation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 323–331, 2019.

[65] F. Noman, S.-H. Salleh, C.-M. Ting, S. B. Samdin, H. Ombao, and H. Hussain, “A markov-switching model approach to heart sound segmentation and classification,” *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 705–716, 2019.

[66] Y. Yin, K. Ma, and M. Liu, “Temporal convolutional network connected with an anti-arrhythmia hidden semi-markov model for heart sound segmentation,” *Applied Sciences*, vol. 10, no. 20, p. 7049, 2020.

[67] C. Liu, D. Springer, and G. D. Clifford, “Performance of an open-source heart sound segmentation algorithm on eight independent databases,” *Physiological measurement*, vol. 38, no. 8, p. 1730, 2017.

[68] F. Renna, J. Oliveira, and M. T. Coimbra, “Deep convolutional neural networks for heart sound segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2435–2445, Nov 2019.

[69] 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, 2016.

[70] Y. Chen, Y. Sun, J. Lv, B. Jia, and X. Huang, “End-to-end heart sound segmentation using deep convolutional recurrent network,” *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 2103–2117, 2021.

[71] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “The physionet computing in cardiology challenge,” 2016. [Online]. Available: https://physionet.org/physiobank/database/challenge/2016/

[72] D. Springer, “Logistic regression-hsmm-based heart sound segmentation. physionet,” 2019. [Online]. Available: https://physionet.org/content/hss/1.0/

[73] Z. Syed, D. Leeds, D. Curtis, F. Nesta, R. A. Levine, and J. Guttag, "A framework for the analysis of acoustical cardiac signals," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 4, pp. 651–662, 2007.

[74] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor, "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results," http://www.peterjbentley.com/heartchallenge/index.html.

[75] Y. Deng and P. J. Bentley, "A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram," in *Workshop Classifying Heart Sounds, La Palma, Canary Islands*, 2012, pp. 1–6.

[76] E. F. Gomes and E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction," in *Workshop Classifying Heart Sounds*, 2012, pp. 480–92.

[77] L. Chun-Lin, "A tutorial of the wavelet transform," *NTUEE, Taiwan*, 2010.

[78] P. K. Jain and A. K. Tiwari, "An adaptive method for shrinking of wavelet coefficients for phonocardiogram denoising," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2016, pp. 1–5.

[79] V. S. Chourasia, A. K. Tiwari, and R. Gangopadhyay, "A novel approach for phonocardiographic signals processing to make possible fetal heart rate evaluations," *Digital Signal Processing*, vol. 30, pp. 165–183, 2014.

[80] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of pcg signals," *IEEE Access*, vol. 8, pp. 160 882–160 890, 2020.

[81] A. Mondal, I. Saxena, H. Tang, and P. Banerjee, "A noise reduction technique based on nonlinear kernel function for heart sound analysis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 775–784, 2017.

[82] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[83] I. Selesnick, "Total variation denoising (an mm algorithm)," *NYU Polytechnic School of Engineering Lecture Notes*, 2012.

[84] T. F. Chan, S. Osher, and J. Shen, "The digital tv filter and nonlinear denoising," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 231–241, Feb 2001.

[85] I. Selesnick, "Total variation denoising (an mm algorithm)," 2011. [Online]. Available: http://eeweb.poly.edu/iselesni/lecture_notes/TVDmm/

[86] M. A. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparisonwith wavalet denoising," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 2633–2636.

[87] T. F. Chan, S. Osher, and J. Shen, "The digital tv filter and nonlinear denoising," *IEEE Transactions on Image processing*, vol. 10, no. 2, pp. 231–241, 2001.

[88] I. W. Selesnick and P.-Y. Chen, "Total variation denoising with overlapping group sparsity," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5696–5700.

[89] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[90] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. Figueiredo, "Adaptive total variation image deblurring: a majorization–minimization approach," *Signal processing*, vol. 89, no. 9, pp. 1683–1693, 2009.

[91] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a gem algorithm exploiting a class of heavy-tailed priors," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 937–951, 2006.

[92] D. Kumar, P. Carvalho, M. Antunes, R. Paiva, and J. Henriques, "Noise detection during heart sound recording using periodicity signatures," *Physiological measurement*, vol. 32, no. 5, p. 599, 2011.

[93] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[94] L. Derczynski, "Complementarity, f-score, and nlp evaluation," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 261–266.

[95] S. Sun, "Segmentation-based adaptive feature extraction combined with mahalanobis distance classification criterion for heart sound diagnostic system," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 11 009–11 022, 2021.

[96] X. Ning and I. W. Selesnick, "Ecg enhancement and qrs detection based on sparse derivatives," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 713–723, 2013.

[97] M. S. Manikandan, K. Soman, and S. Dandapat, "Quality-driven wavelet based pcg signal coding for wireless cardiac patient monitoring," in *Proceedings of the 1st International Conference on Wireless Technologies for Humanitarian Relief*. ACM, 2011, pp. 519–526.

[98] H. M. Al-Angari and A. V. Sahakian, "Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1900–1904, 2007.

[99] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.

[100] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection," *Computers in biology and medicine*, vol. 100, pp. 132–143, 2018.

[101] N. Bacaër, "Verhulst and the logistic equation (1838)," in *A Short History of Mathematical Population Dynamics*. Springer, 2011, pp. 35–39.

[102] J. Berkson, "Application of the logistic function to bio-assay," *Journal of the American Statistical Association*, vol. 39, no. 227, pp. 357–365, 1944.

[103] ——, "Maximum likelihood and minimum x 2 estimates of the logistic function," *Journal of the American statistical association*, vol. 50, no. 269, pp. 130–162, 1955.

[104] L. J. Reed and J. Berkson, "The application of the logistic function to experimental data." *The Journal of Physical Chemistry*, vol. 33, no. 5, pp. 760–779, 1929.

[105] A. P. Kamson, L. N. Sharma, and S. Dandapat, "Multi-centroid diastolic duration distribution based hsmm for heart sound segmentation," *Biomedical Signal Processing and Control*, vol. 48, pp. 265–272, 2019.

[106] I. Rezek and S. J. Roberts, "Envelope extraction via complex homomorphic filtering," *Technical Report TR-98-9, Imperial College, London, U.K.*, 1998.

[107] R. E. Pérez-Guzmán, R. García-Bermúdez, F. Rojas-Ruiz, A. Céspedes-Pérez, and Y. Ojeda-Riquenes, "Evaluation of algorithms for automatic classification of heart sound signals," in *International Conference on Bioinformatics and Biomedical Engineering*. Springer, 2017, pp. 536–545.

[108] D. Van Hoogenhuyze, N. Weinstein, G. J. Martin, J. S. Weiss, J. W. Schaad, X. N. Sahyouni, D. Fintel, W. J. Remme, and D. H. Singer, "Reproducibility and relation to mean heart rate of heart rate variability in normal subjects and in patients with congestive heart failure secondary to coronary artery disease," *The American journal of cardiology*, vol. 68, no. 17, pp. 1668–1676, 1991.

[109] I. O'Brien, P. O'Hare, and R. Corrall, "Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function." *Heart*, vol. 55, no. 4, pp. 348–354, 1986.

[110] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[111] F. Murtagh and P. Legendre, "Ward?s hierarchical agglomerative clustering method: which algorithms implement ward?s criterion?" *Journal of classification*, vol. 31, no. 3, pp. 274–295, 2014.

[112] J. Oliveira, F. Renna, T. Mantadelis, and M. Coimbra, "Adaptive sojourn time hsmm for heart sound segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 642–649, March 2019.

[113] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[114] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[115] P. T. Krishnan, P. Balasubramanian, and S. Umapathy, "Automated heart sound classification system from unsegmented phonocardiogram (pcg) using deep neural network," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 505–515, 2020.

[116] S. Shukla, S. K. Singh, and D. Mitra, "An efficient heart sound segmentation approach using kurtosis and zero frequency filter features," *Biomedical Signal Processing and Control*, vol. 57, p. 101762, 2020.

[117] M. Hamidi, H. Ghassemian, and M. Imani, "Classification of heart sound signal using curve fitting and fractal dimension," *Biomedical Signal Processing and Control*, vol. 39, pp. 351–359, 2018.

[118] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 621–624.

[119] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation," in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 613–616.

[120] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved mfcc features and convolutional recurrent neural networks," *Neural Networks*, vol. 130, pp. 22–32, 2020.

[121] S. Deb and S. Dandapat, "Emotion classification using segmentation of vowel-like and non-vowel-like regions," *IEEE Transactions on Affective Computing*, 2017.

[122] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

[123] 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, 2017.

[124] Y. M. Costa, L. S. Oliveira, and C. N. Silla Jr, "An evaluation of convolutional neural networks for music classification using spectrograms," *Applied soft computing*, vol. 52, pp. 28–38, 2017.

[125] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[126] N. Cristianini, J. Shawe-Taylor, *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

LIST OF PUBLICATIONS

Journal Publications:

1. Alex Paul Kamson, L. N. Sharma, and S. Dandapat. “Multi-centroid diastolic duration distribution based HSMM for heart sound segmentation.” *Biomedical Signal Processing and Control* vol. 48 (2019): 265-272.

2. Alex Paul kamson, L. N. Sharma, and S. Dandapat. “Enhancement of the heart sound envelope using the logistic function amplitude moderation method” *Computer Methods and Programs in Biomedicine* vol. 187 (2020): 105239.

Conference Publications:

1. Alex Paul Kamson, L. N. Sharma, and S. Dandapat. ”Detection of Heart Sound using Logistic Function Amplitude Moderator and Teager-Kaiser Energy Operator.” *2018 International Conference on Signal Processing and Communications (SPCOM)* IEEE, 2018, pp. 422-426.

2. Alex Paul Kamson, L. N. Sharma, and S. Dandapat. “Evaluation of perceptual and multi sub-band energy features for classification of normal, pathological and noisy phono-cardiogram.” *Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC) 2019*.

Workshop attended:

1. As speaker in ”Workshop on Artificial Intelligence in Healthcare Engineering.” *North East Centre for Biological Sciences and Healthcare Engineering (NECBH), IIT Guwahati, (2019)*.