A STUDY ON NOVEL TECHNIQUES FOR HEART SOUND AND MURMUR CLASSIFICATION, SEARCH AND RETRIEVEL

A THESIS

Submitted by

PATIL KIRAN KUMARI

for the award of the degree

of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Dr. M.G.R.
EDUCATIONAL AND RESEARCH INSTITUTE UNIVERSITY
(Declared U/S 3 of the UGC Act 1956)
CHENNAI 600 095

APRIL - 2012

Dr. B. S Nagabhushana
Principal Solution Architect
KPIT Cummins Infosystems Limited
Bangalore.
bsnaga@gmail.com

#708, 9th Cross, 10th Main
II Stage, Indira Nagar
Bangalore 560 038
Mobile: +91-98450 95688

BONAFIDE CERTIFICATE

This is to certify that the thesis titled **"A Study on Novel Techniques for Heart Sound and Murmur Classification, Search and Retrieval"** submitted for the degree of Doctor of Philosophy by Mrs.Patil Kiran Kumari is the record of research work carried out in the Department of Computer Science and Engineering under my guidance and supervision during the period from February 2007 to December 2011 and that this work has not formed the basis for the award of any degree, diploma, associate ship, fellowship or other titles in this University or any other University or other Institution of higher learning.

Signature of the Supervisor
(Dr. B. S Nagabhushana)

Date: 10th April, 2012

DECLARATION

I declare that the thesis entitled **"A Study on Novel Techniques for Heart Sound and Murmur Classification, Search and Retrieval "** submitted by me for the degree of Doctor of Philosophy is the record of research work carried out by me in the Department of Computer Science and Engineering under the guidance and Supervision of Dr. B. S. Nagabhushan during the period from February 2007 to December 2011 and that any part of the thesis has not formed the basis for the award of any degree, diploma, associate ship, fellowship to other titles in this University or any other University or other Institution of Higher Learning.

Signature of the Candidate

Patil Kiran Kumari

Research Scholar
Reg. Number: CSE 07 D011

Department of Computer Science and Engineering

Date: 10th April 2012

DEDICATION

This thesis of mine is

dedicated to my beloved

Parents

who have been my

'Guiding force'

ABSTARCT

Phonocardiography is a specialized domain in cardiology in which the heart sounds and murmurs are recorded as time series data and amenable to the visual inspection as well as audio perception. Digital auscultation – art and science of listening and interpreting heart sounds and murmurs using conventional acoustic or electronic stethoscope and is used as primary screening tool of the heart diseases. The heart sounds and murmurs are the acoustic signals that play significant role in auscultation and clinical interpretations of the heart diseases, in particular heart valvular diseases.

With the recent development in digital signal processing techniques and biomedical instrumentation engineering, it is possible to digitize the heart sounds and murmurs using digital stethoscope and can be stored as digital audio objects in the standard digital audio formats (e.g., .wav or .mp3) in a cardiology database. For example, in Texas Heart Institute, USA, the phonocardiographic recordings (e.g., 300 to 400 recordings/day) are performed and stored in the cardiology database as audio objects. When the phonocardiographic signals are viewed as audio objects, the conventional text processing algorithms for indexing, searching, classification are inappropriate and therefore content – based audio processing algorithms and techniques were proposed.

The content based audio retrieval and classification algorithms are more promising for the multimedia objects; in particular audio objects are used extensively in music information retrieval (MIR) applications. The general procedure is to extract the audio feature vectors and map into the vector space model and apply similarity measures and retrieve the results for an audio query using query by example (QBE).

In this research work, a set of audio content-based retrieval and classification algorithms and techniques exclusively for the heart sounds and murmurs are proposed and investigated. This research contributed to the design of content-based algorithms and novel (psychoacoustic distance measure) techniques for the indexing, searching and classification of heart sounds and murmurs. A new audio model for heart sounds and murmurs is proposed and evaluated at the

signal level (e.g., time series data – characterizing spectral properties) as well as at higher level abstraction – audio perception of heart sound and murmurs. The modeling of audio at frame, inter-frame level along with a set of perceptual features vectors (e.g., pitch, thrill, rush, rumble) and facilitate for the easy retrieval and classification of heart sounds and murmurs. In general, cardiologists listen carefully; evaluate the heart sound and murmurs using subjective reasoning based on audio perceptional properties. A set of novel distance measures based "psychoacoustics" features (e.g., pitch, thrill, soft, loudness etc.) were modeled and quantitatively used for the similarity measures. The psychoacoustics based similarity measures and content based retrieval algorithms were compared with the conventional retrieval (e.g., spectral based) and results establish that the psychoacoustics based retrieval improves the efficiency by 70%.

It is observed that audio retrieval algorithms using psychoacoustic features are in synchronization with the doctor's reasoning with 80 - 85% match when compared with the only spectral properties. On average, it has shown that the MFCC histogram search and retrieval algorithm can achieve a high precision - 97%.

The filtered heart sound signals and murmurs were classified using support vector machine (SVM) technique by varying signal-to-noise ratio (SNR) on reference heart sound and murmur database which contained various types of heart diseases. It is observed that the classification rate is almost 100%. For example, the classification accuracy for diastolic rumble signal is calculated as 320/(320+28+2) = 91.4286%.

Similar results are obtained for other heart sounds and murmurs. Classification plots for different heart sounds and murmurs are reported in the thesis. Discriminatory feature description and an approach for multiclass signal classification based on second-order statistical features are also presented. The 2^{nd} order cumulants of the real and imaginary part of the complex envelope are used as features for multi signal classification. The proposed system is tested on three heart sound and murmur schemes. Two different classifier, Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) classifier are used to classify the heart sound signals. From the results, SVM classifier outperforms the KNN classifier for digital audio signal classification and also it is observed that the classification accuracy of the stenosis scheme for 0 dB is much lesser (i.e. poor) than all other schemes used. Confusion matrix is used to evaluate the performance of the proposed system and the experimental results prove that the proposed system provides satisfactory performance for the multi signal classification.

ACKNOWLEDGEMENTS

While a dissertation is the culmination of an individual student career, it typically benefits from the cherished contributions of many professors, colleagues, family and friends. The greatest contribution has been from my adviser, **Dr.B.S.Nagabhushana**. I have had the pleasure of working with him, learning from him, for many years now. He exhibited a large reservoir of patience and flexibility to allow me to write this Ph. D thesis at my widely varying pace. I thank him for the countless hours he has spent with me, discussing everything from research to career choices, reading my papers, and critiquing my talks. I look forward to knowing him for the rest of my career.

I would like to express my sincere gratitude to **Prof. B.P.Vijaya Kumar**, for his inspiration and support in my research work. Special thanks are given **to Dr. Ravi.S, Dr. Cyril Raj** from **Dr.MGR University and Dr.R.P.Reddy** from REVA ITM Bangalore, for their constant support and advice, which made this journey full of joy.

I was honored to have discussed, and collaborated with several experts **Dr. Ramesh .G**, **Dr.S.B.Patil** and **Dr.H.V.Ramakrishna**. Thanks to them, for the enlightening days. I have been fortunate to have friends like for the moral support and his inspiring words throughout my research work.

My biggest debt is owed to my husband **Mr.Basawaraj Patil**. He had faith in my goals, and always provided the right amount of love and patience particularly, in those many days in which I spent more time with my laptop computer than with him.

Finally, my deepest gratitude is expressed to my family members, especially my brother Mr. Prabhu Dev Patil and all my colleagues & dearest friends at REVA ITM for their comprehensive and selfless support.

Signature of Research Scholar

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMI	Auto Mutual Information
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
AS	Aortic valve Stenosis
AV	Atrio Ventricular
AIR	Audio Information Retrieval
CCI	Cross Correlation Index
COPD	Chronic Obstructive Pulmonary Disease
CBIR	Content Based Information Retrieval
CD	Correlation Dimension
ECG	Electrocardiography signal
EMAT	Electromechanical Activation Time
IR	Information Retrieval
LBNP	Lower Body Negative Pressure
MA	Moving Average
MI	Mitral Insufficiency (Mitral Regurgitation)
MRI	Magnetic Resonance Imaging
PCG	Phonocardiographic signal
PEP	Pre Ejection Period
PPG	Photoplethysmographic Signal
PSD	Power Spectral Density
S1	First Heart Sound
S2	Second Heart Sound

S3	Third Heart Sound
S4	Fourth Sound
SBP	Systolic Blood Pressure
SFFS	Sequential Floating Forward Selection
SVD	Singular Value Decomposition
T1	Recurrence Time of the first kind
T2	Recurrence Time of the second kind
VFD	Variance Fractal Dimension
AS	Aortic Stenosis
MR	Mitral Regurgitation
ESM	Early Systolic Murmur
LSM	Late Systolic Murmur
EC	Ejection Click
DR	Diastolic Rumble
ASD	Atrial Septal Defect
PDA	Patient Ductus Arteriosus
2SS	II Heart Sound Split
3SS	III Heart Sound Split
DSG	Diastolic Summation Gallop
STS	Diastolic Tricuspid Stenosis
SVG	Diastolic Ventricular Gallop
EM	Ejection Murmur

CHAPTER – 1
INTRODUCTION

1.0 INTRODUCTION

Biomedical signals are those signals which are used primarily for extracting information on a biological system under investigation. The process of extracting information could be as simple as feeling the pulse of a person or as complex as analyzing the structure of internal soft tissues by ultrasound scanner or MRI scans. The biomedical signals [R.J. Rangayyan and R.J.Lehner (1988)] originate from variety of sources such as:

- Biomechanical signals based on mechanical function motion, displacement and pressure.
- Biochemical signals are concerned with the biochemical phenomena and chemical composition in blood and serum.
- Bio-optical signals are generated as a result of optical functions.
- Bioelectric signals are the electric fields generated by the action of bioelectric current and voltages. These signals are extensively used in ECG (Electrocardiography) and EEG (Electroencephalography).
- Bioacoustics signals are the acoustics or sound generated due to blood flow and pumping action of heart or flow through heart valves.

The measurement of acoustic signals created by many biomedical phenomena provides information about the underlying system. The examples of such signals are flow of blood in the heart, through the heart's valves and flow of air through the upper and lower airways and lungs which generate heart and lung sounds respectively.

A classical example of acoustic signals is phonocardiography (PCG) signal – an acoustic recording and plot of heart sounds and murmurs signals as function of time. The phonocardiogram is an instrument used for recording the sounds connected with the pumping action of the heart. These sounds provide an indication of the heart rate, rhythmicity, noise-like features, and efficiency of heart valves and provide valuable clinical information. They also give clinical information regarding effectiveness of blood pumping and heart valve action. Auscultation is a process of listening heart sounds and murmurs using acoustical stethoscope and making clinical decisions related to the heart diseases, in particular the valvular diseases. The phonocardiography (PCG) provides a recording of the heart sounds. This information is diagnostically more important. Many diseases of the heart cause changes in the heart sounds and additional murmurs before other signs and symptoms appear. Hence heart sound analysis by auscultation is the primary test conducted by physician or cardiologists to assess the condition of the heart.

PCG signals or heart sounds have been studied [D.H.Bekkering, Weber J. (1957)] from past many years. Phonocardiography plays an important role in cardiac care as they are non-invasive, non-expensive but accurate monitoring method for valves functioning and easily repeatable with no risk to the patient. However, heart diagnosis by auscultation requires high skills and experience of the listener [A.A Luisada, (1965)]. Heart failure and stroke cause big burden on society due to their high costs of care, lower quality of life and premature death.

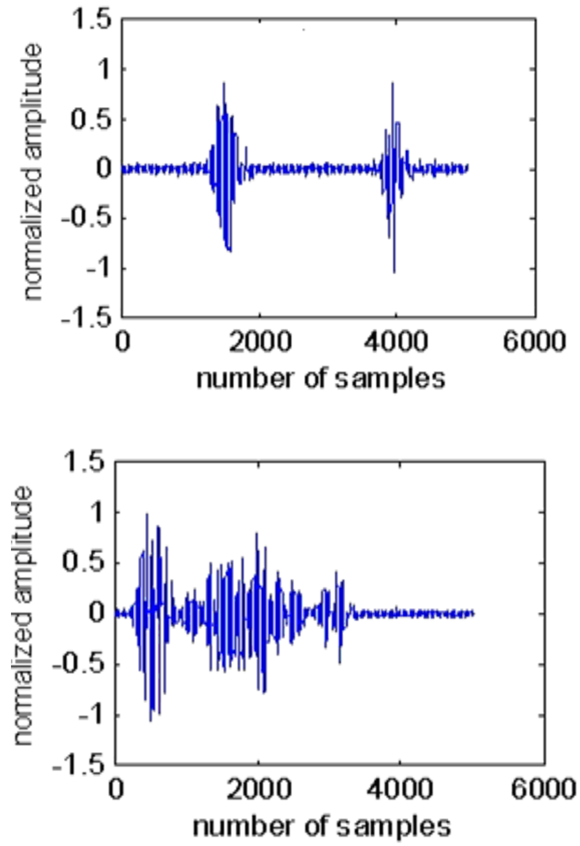
Technological advances have been facilitating the research into both the creation of new areas and the development of existing methods for monitoring physiological signals. The application of engineering to this biomedical problem is appropriate, as scientific measurement theory is well in advance of technology used in clinical situations. The specific area which this thesis addresses is the sound or acoustic signals produced by the heart, their retrieval from the PCG audio database and classification of cardiovascular diseases. In particular, pathological conditions of the heart produce sounds which are different from those of the "normal" heart. As such, the transduction of these sound vibrations may be used for the detection and classification of heart pathologies. Previous efforts [H.Liang, S.Lukkarinen and Hartimo (1999), H. Ljunggren (1949)] in the area of heart sound digital signal processing have been pursued which provided a background for this work.

1.1 HEART SOUNDS AND MURMURS

The PCG data acquisition, data storage and file format, transmission and interpretations of the phonocardiography signals are technically challenging and used in the detection and investigation of heart diseases such as stenosis, heart valve insufficiency, septal defect and patent ductusis. The audio signals of heart diseases and murmurs have various spectral properties (e.g., magnitude spectrum, centroid, zero crossing etc.) and have been studied extensively using digital signals processing (DSP) techniques. However, the heart sounds and murmurs posses another dimension based on the audio perception properties (e.g. pitch, intensity, loudness, noise-like, thrill, gallop etc.). The audio perception of the heart sounds play significant role in auscultation and doctors interpret the heart diseases and murmurs based on listening of heart sounds and murmurs and helps in clinical diagnosis [Erickson (1997), Gutyon (2005)].

The typical clinical procedure is to lay down the patient in supine position and perform silent respiration in a quite clinical setting. The digital or electronic stethoscope is placed in the mitral area. The digitally processed waveform records the PCG signal characteristics, time intervals measured in milliseconds and thus obtain the mitral component (from Q in QRS complex to first heart sound -S1), cardiologic systole (S1–first sound to S2–second heart sound), heart rate, electromechanical systole, cardiologic systole index etc.. These are very useful to derive the heart and valve parameters and detection of various cardiovascular diseases [Mannheimer (1956)].

A normal cardiac cycle contains two major sounds: first heart sound (S1) and second heart sound (S2). The S1 occurs during isovolumic contraction period consisting of mitral valve closure, tricuspid valve and aortic valve opening in a specific order. When the ventricular pressure exceeds the aortic pressure, the aortic valve opens, the pressure in the ventricle falls below the aortic pressure and the aortic valve closes giving rise to S2. The pressure in the ventricle drops steeply giving rise to S3. Similarly, the atrial systole may also produce the audible forth heart sound S4 [Michael (2002)]. These heart sounds (S1, S2, S3 and S4) carry clinical information and used in clinical diagnosis. Murmurs are noise-like sounds caused by the cardiovascular diseases, aberrations and defects in heart valve functioning. A PCG signal has unique characteristics such as low frequency, period and duration of the occurrence of heart sounds, noise, signal amplitudes and has close relationship with the electrocardiograph (ECG) and contains valuable clinical information. The analysis of the PCG helps in detecting various types of heart diseases; in particular heart valve diseases and accurate calculation of heart functioning, volume and pumping efficiency of the heart valves. A normal and abnormal cardiac sounds profile due to aortic stenosis is shown in Figure 1.1.



(a) Normal cardiac sound (b) Abnormal cardiac sound

Figure 1.1 Normal and abnormal cardiac (aortic stenosis) PCG (adopted from R.J. Rangayyan, R.J.Lehner (1988)).

Auscultation is a clinical procedure of listening and interpreting heart sounds and murmurs [T. Michael (1997)]. However, the heart diagnosis by auscultation requires special skills and professional experience to make correct interpretation and clinical diagnosis [Baidt (2006)]. With recent advancements in biomedical signal processing, the digital phonocardiography offers powerful techniques and methods by applying the advanced signal processing techniques that can be used for classification and retrieval applications of heart sounds and murmurs and helps in clinical diagnosis.

1.2 PHONOCARDIOGRAPH SIGNALS AND ACOUSTIC PROPERTIES

For the past few decades, the phonocardiograph signals or heart sounds have been studied extensively [Bekkering and Weber (1957), Leatham (1975), Erickson (1997)] and plays an important role in investigations and detection of cardiac diseases. The phonocardiography is a non-invasive, low-cost investigation method used for analyzing heart sounds to obtain the clinical information about the heart disease; in particular heart valve diseases. The genesis of the heart sounds is well established and constitutes a separate study in cardiology [Chad Stoltz, Robert J. Bryg, (2002) and B.G. Wells, Rapport M.B., Sprague H.B. (1949)]. A phonocardiograph is an instrument used for the recording the cardiac sounds and printed on a thermal paper or displayed on the monitor. From the beginning till today, the principal instrument used for the clinical investigation of heart sounds is the acoustical stethoscope.

With recent developments in digital technology and signal processing, electronic or digital stethoscope (e.g., Philips 5003X or Meditron 3412) are common and extensively used in phonocardiology. A typical digital stethoscope consists of high sensitivity microphones (contact or air coupled or crystal type), signal amplifiers with bandwidth in a frequency range of 20 - 2000 Hz, and filters (high pass, low-pass, and notch) for specific selection of the PCG signals, analog-to-digital converter – ADC (16 bit), input-output interfaces for PC and Internet connectivity. It is also possible to store the heart and murmurs sounds as audio objects in the databases using the standard digital audio .mp3 or .wav format.

Table 1.1 Properties of heart sounds.

Sound	Location (ms)	Duration (ms)	Frequency Range (Hz)
S1	10 – 50 after R-peak in ECG	100 – 160	10 – 140
S2	280 – 360 after R-peak in ECG	80 – 140	10 – 400
S3	440 – 460 after R-peak in ECG or 120 – 180 after closure of semilunar valves	40 – 80	15 – 60
S4	40 – 120 after beginning of P-wave in ECG	30 – 60	15 – 45

The acquisition and storage of heart sounds and murmurs form the pre-processing steps and this research work is mainly concerned with the processing (retrieval and classification) of heart sounds and murmurs from the acoustic or perceptual perspective. Table 1.1 describes the specific properties such as location with respect to ECG, duration and frequency of (S1, S2, S3, and S4) heart sounds. The timing of the S1, S2, S3, and S4, with respect to ECG cycle and sequence of opening and closing of heart valves is described in Figure 1.2.

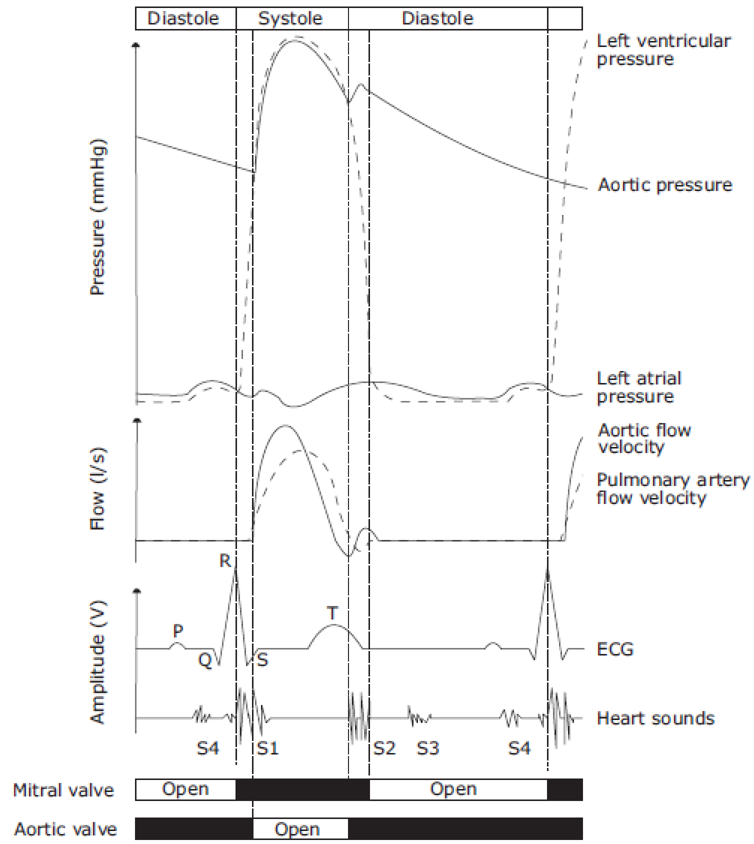


Figure 1.2 Heart sounds (S1, S2, S3, and S4) and ECG signal.

1.3 ISSUES AND CHALLENGES IN HEART SOUNDS AND MURMURS RETRIEVAL AND CLASSIFICATION

The multimedia data consisting of text, audio, images and video for classification and retrieval applications have different challenges when compared with text retrieval application [Kemp (1995), Wold (1996)]. Increasing the number and types of multimedia data (text, audio, video, and images) requires a reliable data management system in terms of storage, retrieval, and dissemination of multimedia data. The techniques of text classification and retrieval algorithms are inadequate for classification and retrieval the multimedia objects. Based on the state of the art of content-based multimedia information retrieval [A Yoshitaka and T. Ichikawa (1999)], the main issue in multimedia information retrieval is how to bridge the semantic gap when text annotations are nonexistent or incomplete. A potential research area is how to develop the direct access method into an audio object to find knowledge or meaningful information that can improve retrieval accuracy and speed up the process even when text annotations are present by giving additional insight into the media collections and have to use the content-based systems information retrieval techniques.

The audio retrieval and classification problems are studied in [Wold (1996), Li and Li (2003), Foote (1997)] Audio Information Retrieval (AIR) [Rolland et el., (1999)] and Music Information Retrieval (MIR) [Uitdenbogerd and Zobel, (1999), Ghias et el., (1995)] systems and related research topics [Christopher (2005), Katz, Burkard and Medwetsky (2002)]. In particular, the research in speech signals processing [Rabiner and Schafer, (1978)] and speech recognition processing [Rabiner and Juang, (1993)] contribute to the audio processing. The audio objects (e.g., songs) are modeled as feature vectors and employ the similarity measures for speech or music retrieval [Raghvan and Gudivada (1995)].

The proposed research work extends these content-based retrieval and classification techniques exclusively for heart sounds and murmurs. The phonocardiographic signal is an important biomedical signal of the heart of audio nature and related to the contractile activity of the cardiohemic system and represents a recording of the heart sound signal and considers them as audio objects. The proposed research work explores the possibility of audio modeling of heart sounds and murmurs based on spectral and perceptual properties. Once the audio objects are in the heart and murmurs database, the next challenge is to classify them and retrieve the matched audio objects from the audio query as posed by the doctor using query by example (QBE) paradigm [Rolland, Raskinis and Ganascia (1999)]. When the cardiologists interact with the heart and murmur database and query these databases consisting of audio recording of the heart sound and murmurs, the retrieval and classification of heart sounds and murmurs becomes more challenging when compared with conventional text retrieval. The research in the text search and retrieval is extensively studied and these methods are not appropriate for the audio objects – in particular, classification and retrieval tasks of audio objects [Downie (2006), Foote (1999)].

1.4 RECENT WORK IN CONTENT BASED INFORMATION RETRIEVAL FOR AUDIO OBJECTS

The audio retrieval research is mainly concerned with the classification and retrieval applications and notion of audio objects is broad and mainly falls into the three categories: speech, music and natural sounds. The speech processing and recognition is well studied [Rabiner, L. and Juang, B (1993)] and successfully applied in various applications such as speaker recognition, speaker identification and speech based user interfaces and interactions.

Music is also a classic example of audio objects. The music research [Martin, K. (1998), L. Kjell and L. Pauli (1998)] is mainly concerned with the analysis of music, classification music, and recognition of music genera, classification of musical notes, music recommendation systems and content based applications.

Recently, there has been interest [D. Mitrovic and M. Zeppelzauer. (2006)] in the classification of natural sounds such as those of birds and other animals. The focus is to perform a matching of the unknown audio signature and retrieve the matched natural sounds.

In this research work, heart sound and murmurs are modeled as audio objects and scope of the subsequent discussions is confined to the modeling of heart sounds and murmurs as audio objects, heart sound modeling using psychoacoustic framework, classification and retrieval of heart sound and murmurs. The main objective of this research is concerned with the indexing, searching, retrieval and classification of heart sounds and murmurs. It also discusses the design, implementation and evaluation of various searching and classification algorithms in a clinical setting and use test samples from the heart sounds and murmurs databases.

In the following discussion, various content based algorithms from literature and discus their contributions. Foote [J. Foote (1999)] presented an idea for the representation of an audio object by a template that characterizes the object in his purposed system. For construction of a template; an audio signal is first divided into overlapping frames of constant length then using simple signal processing techniques, for each frame a 13-dimensional feature vector is extracted (12 Mel-Frequency Cepstral Coefficients – MFCC plus 1 energy) at a 500 Hz, and then these feature vectors are used to generate templates using tree-based Vector Quantizer trained to maximize mutual information (MMI). For retrieval, the query is first converted into a template and compared with the MFCCs of the audio object in the database using similarity distance function and finally a ranked list is generated based on minimum distance. In this, performance of the system with euclidean distance as well as cosine distance is also compared, and experimental results show that cosine distance performs slightly batter than Euclidean distance. This system may fail for music retrieval if either query is corrupted with noise or bad quality of music is recorded. These generic algorithms are not appropriate for heart and murmurs sounds as they are non-stationary and also contain noisy data.

A commercial company - Muscle fish group [*http://www.musclefish.com*], in this system an audio object is characterized by its frame level and global acoustical and perceptual parameters. These features are extracted at frame level using signal processing techniques and globally using statistical analysis based on frame level features and musical features (for music signals only) using musical analysis. Frame level features consist of loudness, pitch, tone (brightness and bandwidth), MFCCs and their derivatives. Global features are determined by applying statistical modeling techniques on the frame level features that is, using Gaussian and Histogram Modeling (HM) techniques to analyze audio objects. For musical objects, musical features (i.e. rhythm, events and distance (interval)) are extracted using signal processing techniques like pitch tracking, voiced and unvoiced segmentation and note formation. The authors [Blum, T., Keislar, D., Wheaton, J., and Wold, E. (1997)] use multidimensional features for indexing, distance measure for retrieval. These approaches improve the performance, and a modified version of query-point-expansion technique is used, and works for QBE only. This work closely matches the present work however, the audio objects considered are restricted to music content and classification accuracy is compromised and hence may not be acceptable to heart and murmur sounds.

Authors [R. Subramanya and A. Youseef (1998)], represent an audio object by 3-dimensional feature vector for their system. These are standard DSP techniques and based on spectral properties. Feature vectors are extracted at frame level using Discrete Wavelet Transform (DWT). They applied 4-level DWT decomposition to audio signal, then transformed into domain variance, zero-crossing rate and mean of wavelet coefficients are determined to form feature vectors. For indexing structure, B-tree structure [D. Comer (1979)] is used, which is constructed using clustering technique along with multi resolution property of the wavelet transform. Similarity search is applied using weighted Euclidean distance, and based on minimum distance a ranked target list is retrieved for the desired query. First of all an audio signal is decomposed using 4-level DWT and then form wavelet coefficients, feature vector is formed using all approximate coefficients and 20% - 30% of detail coefficients of all levels obtained during wavelet decomposition. For query processing same procedure is applied. They did not specified any indexing technique but for similarity matching they used Euclidean distance measure. Detailed discussion on retrieval strategies currently employed in the Information Retrieval (IR) area can be found [D. A. Grossman and O. Frieder (1998)]. Previous research work carried out in the area of Content Based Information Retrieval (CBIR) motivated us to explore the applicability of CBIR techniques for classification and retrieval of biomedical signal, in particular heart sounds and murmurs.

In brief, the proposed research work models the heart sounds and murmur as audio object for representation using spectral and psychoacoustic properties, apply various similarity based indexing, searching and classification of heart sounds and murmurs, study the applicability of various audio retrieval algorithms under optimal conditions, performance evaluation of algorithms (in terms of recall and precision).

1.5 MOTIVATION

Texas Heart Centre, USA [www.texasheart.org] uses extensively electronic phonocardiography using electronic stethoscope as a preliminary diagnostic tool for cardiovascular disease and murmur detection. On an average, doctors acquire and process audio recording of heart sound murmurs to about 300 - 500 recordings per day. The duration of each audio recording is about 1 min. to 10 min. in a hospital environment. The doctors have collected huge audio data (i.e. heart sounds and murmurs) and upload into central server and integrated with the hospital and patient information systems. When the doctors want to compare the new PCG recordings of unknown cardiovascular diseases or murmurs, it is critically necessary to classify, search and retrieve the heart sounds that match with the known heart disease and murmurs from the database based on content based classification and retrieval techniques. The doctors prefer to query the audio objects by a technique called query by example (QBE), in which the doctor submit an audio query i.e. the recording of the heart sound and murmurs and expect the system to match the best query based on the search or query criteria and derive the matched heart sounds and murmurs from the heart sound databases. Hence, we need a framework that support the heart sound modeling, content-based, similarity-based algorithms which classify, match, index and retrieve the best matched results with high relevance and assist the doctor for better clinical interpretations of the heart disease and murmurs. In Ireland, University of Dundee [http://www.dundee.ac.uk] has collected huge database of heart sounds and murmurs from the community health care center. With the recent development in Information and Communication Technology (ICT) and telecardiology applications, the PCG collection, storage and transmission based on health standards are emerging and also need for the efficient classification algorithms and retrieval systems that assist the doctors for effective clinical investigation such as Decision Support System (DSS).

According to WHO survey [http://www.who.int] with specific reference to India, it is found that there is not sufficient number of cardiac experts for community services specific to the cardiovascular disease. With ICT enabled healthcare systems, the physiological signals including, PCG can be integrated with patient health care system and PCG will be used as a primary screening tool. When the acquired data (physiological signals - ECG and PCG) is processed in the patient information systems, there is a critical need for a system which will help analyze the PCG data quickly by paramedics who are not experts in cardiology. The acquired and processed PCG data can be sent to the cardiologist using ICT enabled infrastructure (e.g., internet or mobile) and assist cardiologists to arrive at better clinical interpretation using both quantitative and qualitative techniques (digital auscultation using .mp3 recording and visual inspection) and retrieve the best matched heart disease query from the heart database. It is aimed to address one of such community healthcare problem.

Looking at the need for remote quality and immediate medical care where there are few doctors in the rural parts of India, we found that remotely acquiring and diagnosing patients by experts will address the problem. Efficient classification and retrieval algorithms would be key contribution in the area of telemedicine, in particular telecardiology [Stein (1981)]. The acquisition, storage, transmission and interpretation of PCG (heart sounds and murmurs) signals are using different formats, incompatible structures, and propriety protocols creating serious interoperability issues in health care and tele-cardiology applications in particular classification and retrieval problems. These application scenarios in cardiac healthcare motivated to take up the challenges of efficient classification and retrieval of heart sounds and murmurs.

1.6 PROBLEM STATEMENT AND PROPOSED SOLUTION

Heart sounds are caused by turbulence in blood flow and vibration of cardiac and vascular structures. In this work, heart sounds and murmurs are processed and the spectral and perceptual features are extracted and used for the classification and retrieval applications. The block diagram of the proposed system is show in Figure 1.3. The dimension of the feature space is set equal to the number of extracted features.

Before any audio signal (heart sounds) can be classified under a given class, the features in that signal are to be extracted. In this work, the specific heart sounds and murmurs are studied as the audio signal produced by the heart. The heart sounds and murmurs are visualized as a sound or audio signal contains much useful information and can assist the doctor in diagnoses of the heart diseases. Specifically, it is proposed that much information is derived by audio content based processing techniques.

In this work, from the input heart sounds, two measures are used and different features are extracted. Normal and various types of abnormal heart sound signals were studied and the features (magnitude and power spectrum, FFT, Zero Crossing, MFCC etc.) were extracted from each case. The audio similarity measures plays significant role in the indexing and retrieval of matched/classified hearts sounds and murmurs form the heart sound database. This research work explores various clustering techniques (K-means, Hierarchical etc), histogram matching algorithms, distance functions (Euclidian distance, Hamming distance, cosine etc.) are investigated and their applicability for the heart sounds and murmurs classification and retrieval for indexing and fast retrieval from the large heart sound and murmurs databases of audio recordings. It also explores the performance of these algorithms in terms of relevance, precision and recall.

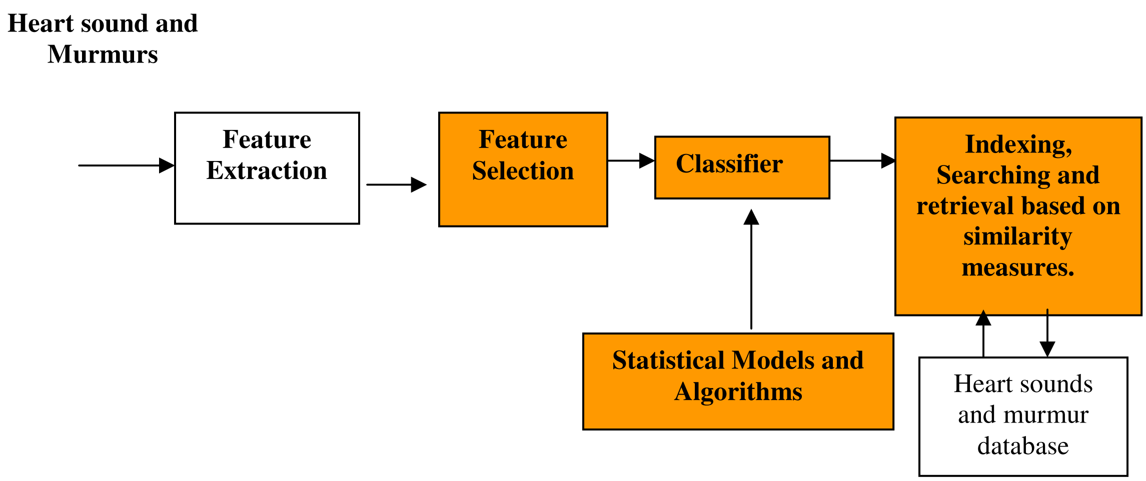


Figure 1.3 Block diagram of the proposed work.

In the proposed system, the feature vectors (spectral and perceptual) are used and support the Query by Example (QBE) in which the doctor can submit the recorded heart sound and query the heart sounds and murmur database. The proposed system will "find" the distance function which has minimum distance function of various distance function types (e.g., Euclidian) and they are indexed and ranked on the bases of psychoacoustics features of the query and reference heart sounds and murmurs. Typical spectral features such as energy in S1, energy in S2, murmurs exceeding 200 Hz frequency, first frequency peak, murmur energy ratio, sample entropy, auto mutual information are investigated in this research work. Apart from that, this work propose a novel psychoacoustics framework for audio modeling and retrieval based on psychoacoustic properties of the query (e.g., pitch, loudness, thrill, and gallop) as normally used by the doctor rather than the spectral properties. A typical QBE query from a doctor perspective could be "*give me all the heart and murmur sound which has high pitch and loudness in S1?*" In brief, the scope of the work is confined to the classification, indexing, searching and retrieval of heart sounds and murmurs for a given audio query, derive best similarity – based algorithms used for classification and retrieval applications of heart diseases. In this research, these issues are focused and acceptable results are obtained.

1.7 PSYCHOACOUSTIC MODELING OF HEART SOUNDS AND MURMURS

Cardiac auscultation is highly subjective and a cognitive process and the amount of information that can be obtained by listening heart sounds largely depend on the expertise, and experience of the physician. In general, the classification and interpretation of heart sounds and murmurs is based on a adjective 0-6/6 grade scale and described by using “faint”, “soft”, “loud”, “ high pitch”, “clear”, “thrill”, “tremor”, “musical” and others terms [Leung, White, Collias, Brown and Salmon (1982)]. These terms are not well-defined and suitable mathematical models are not available. Apart from that, the adjective scales vary among the doctors and difficult to derive a standard model for heart sound quality and correct clinical interpretation. This research work proposes a novel framework based on psychoacoustic principles and derive psychoacoustic models for the heart sounds and murmurs. It also discuss the theoretical foundations, psychoacoustic principles and derive the mathematical models for the psychoacoustic features such as loudness, sharpness, intensity, strength, roughness, tonality etc. for a set of heart sounds and murmurs [Stevens, Volkmann and Newman(1937), Ristama (1967)]. The proposed framework helps in deriving heart sound quality and also used for the comparison and correlation with normal and pathologic murmurs and enhances clinical decisions. When analyzing human heart sounds, there is information which may be gained by simply listening to the heart sounds as in cardiac auscultation, and by seeing the heart sound waveform as in phonocardiography [Shipton and Wahba (2001)]. However, more extensive analysis using psychoacoustic modeling of heart sounds some basic signal processing methods may provide further insight into the state of the heart. This research work attempts to take advantage of the content-based, psychoacoustic features for the accurate classification and efficient retrieval of heart sound and murmurs.

1.8 CONTENT BASED RETRIEVAL OF PCG SIGNALS BASED ON PSYCHOACOUSTIC PRINCIPLES

Auscultation is quite reliable technique at the same time quite difficult to master. With the stethoscope, it requires highly experienced professionals to read and understand phonocardiograph. Content-based indexing of audio (and multimedia) data has become more important since conventional databases cannot provide the necessary efficiency and performance [E. Wold et al., (1996)]. However, there are few main difficult problems. First, the content of audio data is subjective information; it is hard to give the descriptions in words. The recognition of data content requires prior knowledge and special techniques in Signal Processing and Pattern Recognition, which usually require long computing time. Second, since several audio features can be used as indices [Eronen, A., Klapuri, (2000)] (such as pitch, amplitude, and frequency), a method or processing technique designed and developed for one feature may not be appropriate for another. Thus, there arises the necessity of content based analysis of phonocardiography (PCG) signal. This research work proposes a PCG retrieval system, based on modular architecture consisting of feature extraction, histogram modeling, pattern matching using different content-based audio processing techniques and retrieval algorithms.

1.9 OBJECTIVES OF THE WORK

- To design and implement a framework for audio modeling of heart sounds and murmurs using feature vectors (spectral, and perceptual) and implementation of content based heart sound and murmurs retrieval algorithms and auditory user interfaces for cardiologist, in which doctor can directly submit audio query and obtain the ranked heart and murmur audio files using similarity measures. It also includes comparative studies and recommends best retrieval algorithms for heart sound retrieval applications.

- To derive psychoacoustic models and characterize psychoacoustic features such as loudness, pitch, sound, intensity, sharpness and spectral properties (centroid, zero crossing etc) that help efficient retrieval of heart sounds for a given reference or audio heart sounds.

• Design and implementation of set of algorithms using classifier techniques such as Support Vector Machines, K-Means, Histogram Matching, similarity measures, and distance function are used for audio retrieval of heart sounds from heart sound and murmur data base and validate with experimental results.

2.0 ORGANIZATION OF THE THESIS

Chapter 1 describes the introductory concepts of phonocardiogram signal, their advantages in clinical decision analysis along with the problem statement and proposed solution. The objectives of the research, methodology and the benefits are also presented in this chapter.

Chapter 2 describes the literature review on the relevant works done so far in the chosen area of research. Here, a literature survey is presented from two perspectives: the digital signal processing perspective and content based audio retrieval algorithms and applications.

Chapter 3 describes the preliminary features associated with the heart sound, origin of heart sounds and murmurs with clinical perspective, DSP signal and the techniques that are to be used to extract the features, properties of spectral and perceptual understanding of heart sounds and murmurs and their clinical significance is highlighted.

In chapter 4, the novel implementation of hierarchical clustering and retrieval algorithms are discussed and derive optimal conditions for the best retrieval performance**.**

Chapter 5 presents the psychoacoustic modeling of heart sounds and audio presentations of feature vectors are discussed based on the psychoacoustic framework. Different heart sound murmurs corresponding to both normal and abnormal subjects taken in real-time are also discussed in detail. These PCG signals are analyzed and the different features are extracted. The algorithms are implemented using the MATLAB simulator and the extracted features are reported which posses discriminatory properties.

Chapter 6 presents the audio object information retrieval, indexing and searching along with experiments and results. A formal model of the indexing and searching is also presented.

Chapter 7 presents the design of classifier network for PCG signal with the extracted features is presented. The Support Vector Machine (SVM) classifier with indexing and retrieval of heart sounds based on distance measures are implemented using the MATLAB and the classifier output is presented. Content based PCG retrieval techniques along with experimental results are presented.

Chapter 8 presents the research findings and summary on inference from results and future research directions.

CHAPTER – 2

LITERATURE REVIEW

2.0 INTRODUCTION

In this chapter, a literature review of PCG signals, from time domain, frequency domain, and time-frequency domain signal processing perspective is presented. Additionally, a review on audio classification and retrieval methodologies is discussed. Literature review is divided into two major parts based on the DSP and psychoacoustic perspective. The first part is concerned with bio-signal and PCG, signal processing perspective using various standard techniques [Michael H. Crawford, MD, 2002 & M. Carey, E. Parris, and H. Lloyd-Thomas, 1999] such as FFT, time-scale map obtained from a proposed wavelet based approach, automatic classification of the phonocardiogram are discussed. The second part reflects the review done on concepts of psychoacoustic modeling [E. Zwicker and H. Fastl, (1999)] and state of art review on content based audio retrieval techniques and methodologies [Kemp (1996), Foote (1997), Tzanetakis and Cook (2000)].

2.1 PHONOCARDIOGRAPHY SIGNAL PROCESSING: MEDICAL PERSPECTIVE

In [Michael H. and Crawford, (2002)], the patient's history was a critical feature in the evaluation of suspected heart disease. It includes information about the present illness, past diseases, and the patient's family medical history. A frequent complaint of patients with a variety of cardiac diseases, for example dyspnea may have many causes. It was therefore useful to search for a history of rheumatic fever, which may manifest as Sydenham's chorea, joint pain and swelling, or merely frequent sore throats. Proper measurement of the systemic arterial pressure by cuff sphygmomanometer was quoted as one of the keystones of the cardiovascular physical examination. Assessment of the jugular venous pulse can provide information about the central venous pressure and right-heart function. Examination of the right internal jugular vein was ideal for assessing central venous pressure because it was attached directly to the superior vena cava without intervening valves. The first heart sound was coincident with mitral and tricuspid valve closure and has two components in up to 40% of normal individuals. There was little change reported in the intensity of this sound with respiration or position.

In [Peter C. Chien and William H. Frishman, (2002)], a great deal of emphasis has been placed on the relationship between elevated serum cholesterol levels, especially low-density lipoprotein cholesterol (LDL-C) and the incidence of coronary artery disease (CAD). The exogenous pathway was mainly responsible for absorption of dietary fat in the postprandial state and its subsequent distribution to the tissues. Current concepts in atherosclerosis suggest that oxidation of LDL was involved in its pathogenesis. A reliable direct method for measuring low-density lipoprotein cholesterol low-density lipoprotein cholesterol was needed because the accuracy of indirect estimates of LDL-C reflects measurements of total-C, and triglycerides, each of which contributes some degree of imprecision. In [Michael H. and Crawford, MD, 2002], the clinical purposes, patients with chronic ischemic heart disease fall into two general categories: those with symptoms related to the disease, and those who are asymptomatic. Myocardial ischemia was the result of an imbalance between myocardial oxygen supply and demand. Coronary atherosclerosis and other diseases reduce the supply of oxygenated blood by obstructing the coronary arteries. The physical examination was often not helpful in the diagnosis of chronic ischemic heart disease. This was because many patients with chronic ischemic heart disease have no physical findings related to the disease, or if they do, the findings are not specific for coronary artery disease. This can be done with ambulatory electrocardiogram (ECG). Under unusual circumstances, a patient may have spontaneous angina or ischemia in a medical facility, where it was possible to inject a radionuclide agent and immediately image the myocardium for perfusion defects.

In [Prediman K. Shah and Kuang-Yuh Chyu, (2002)], the term intermediate coronary syndrome was used to describe what was now known as the syndrome of unstable angina, which has also been given numerous other labels: preinfarction angina, status anginosus, crescendo angina, impending myocardial infarction (MI), coronary failure, acute coronary insufficiency, spasmodic angina, and atypical angina. It was clear that coronary artery disease, the primary cause of mortality and morbidity in much of the industrialized world, takes its toll through such acute complications (unstable coronary syndromes) as unstable angina, myocardial infarction, acute congestive heart failure, and sudden cardiac death. Although the precise mechanisms are not known, several hypotheses explain the propensity of plaques to rupture. These include circumferential hemodynamic stresses related to arterial pulse and pressure, intraplaque hemorrhage from small intimal fissures, vasoconstriction, and the twisting and bending of arteries.

In [Allan S. Jaffe and Wayne L. Miller (2002)], acute myocardial infarction (AMI) is a clinical syndrome that results from an injury to myocardial tissue that is caused by an imbalance between myocardial oxygen supply and demand. The occlusion is thought to develop in response to plaque rupture when the luminal diameter of the coronary artery is sufficiently reduced to initiate clot formation or if erosion of the plaque causes exposure of procoagulant factors. In [Craig Timm (2002)], the initial development of coronary care units and rapid cardio version or defibrillation of life-threatening ventricular arrhythmias followed by risk-factor modification and such major advances as thrombolytic therapy and emergency revascularization, have contributed significantly to the successful care of the acute myocardial infarction patient. The path physiology of cardiogenic shock in acute infarction complicated by mechanical problems is somewhat different. It may be followed by chest pain, but they more commonly present abruptly as acute pulmonary edema or cardiac arrest.

In [Kathleen M. and Allen (2002)], a noninvasive or invasive assessment, generally driven by a clinical syndrome, leads to a diagnosis of clinically significant coronary artery diseases. The use of cardiac catheters quickly advanced, and many discoveries were made that enhanced our understanding of coronary path physiology. Balloon catheters were bulky and difficult to maneuver. A number of clinical trials have shown that the use of IIb/IIIa inhibitors has reduced acute closure and recurrent events by as much as 30% in a broad spectrum of patients. The patient receives a continuous infusion or a “cocktail” consisting of calcium channel blockers and nitrates to prevent vasospasm during the rotational atherectomy. The popularity of stents exploded when the “stress” and “bent-stent” trials were published. A laser source is transmitted through a fiber optic cable to the end of an angioplasty balloon. The energy is transmitted from the fiber to the plaque, thereby melting and ablating the plaque.

In [Blase A. Carabello and Michael H. Crawford (2002)], the concentric left ventricular hypertrophy that develops as a major compensatory mechanism helps the left ventricle cope with the increased pressure work it must perform. The onset of these symptoms heralds a dramatic increase in the mortality rate for these patients if aortic valve replacement is not performed. The left ventricular myocardium must produce stress in order to shorten, maintaining normal stress facilitates shortening. The increased diastolic filling pressure is referred to the left atrium and to the pulmonary veins, where pulmonary venous congestion develops, leading to increased lung water, increased lung stiffness, and dyspnea. A palpable shuddering sensation of the carotid pulse may also be noted. In elderly patients, increased stiffness of the carotid arteries may falsely normalize the upstroke, making it feel relatively brisk in nature. The presence and severity of coronary disease will influence the course of therapy, tipping the balance toward surgery when the aortic stenosis is of borderline severity. It should be noted, however, that when prosthetic stenos is suspected, the gradient across the valve may be difficult to obtain invasively.

In [Michael. C (2002)] the presentation and findings in patients with aortic regugration (AR) depend on its severity and rapidity of onset. The acute ventricular volume overload therefore results in a small increase in end-diastolic volume and severe elevation of end-diastolic pressure, which is transmitted to the left atrium and pulmonary veins, culminating in acute pulmonary edema. Arrhythmias, including ventricular ectopy and ventricular tachycardia, may occur in advanced cases with left ventricular dysfunction. Echocardiography imaging with M-mode and two-dimensional (2D) examinations cannot detect the presence of aortic insufficiency but can provide indirect clues to the presence of aortic insufficiency. Similarly, a good estimation of AR severity has been found by relating the cross-sectional area of the jet at its origin to the left ventricular outflow area. The availability of other independent Doppler indices of AR severity further allows the corroboration of color Doppler findings. Regurgitant fraction is derived as the regurgitant volume divided by left ventricular stroke volume. Presently it is recommended that all patients who have taken any of these drugs should have a clinical examination.

In [Chad Stoltz and Robert J. Bryg (2002)], the normal mitral apparatus is a complex structure whose components must permit a large volume of blood to pass from the left atrium to the left ventricle. As the stenosis progresses, a transmitral pressure gradient develops to facilitate flow across the stenotic valve in diastole. The disproportional increase is termed reactive pulmonary hypertension. The vigorous opening of the leaflets is secondary to high left a trial pressures accompanied by a fall in left ventricular pressures in early diastole. Radiologic examination of the lung fields reveals elevated pulmonary pressures. The pressure half-time is the time required for the peak pressure gradient between the left atrium and the left ventricle to decline to one half of its original value. The natural history of mitral stenosis has been profoundly influenced by the advancement of cardiovascular interventions. In most patients, rheumatic mitral stenosis a progressive disease. The procedure should therefore be performed prior to the development of significant left ventricular dysfunction.

In [Michael H. and Crawford (2002)], anything that causes left ventricular dilatation may disrupt the alignment of the papillary muscles, impairing their function and dilating the annulus, resulting in mitral regurgitation. An increase in the middle connective tissue layer of the mitral valve causes an increase in leaflet size and elongated chordate. As the severity of mitral regurgitation increases over time, the ability of the dilated left ventricle to augment systolic function reaches its limits, left ventricular systolic function falls, and heart failure ensues. Occasionally, in patients with posterior leaflet defects, the direction of the regurgitant jet may be anterior, and the murmur is heard in the aortic area. Electrocardiography (ECG) exercise testing is usually done only to confirm the patient's physical limitations, because ECG changes in the face of a left ventricular volume load are not likely to be accurate for the diagnosis of coronary artery disease. This assessment method correlates well with the color-flow Doppler system. In addition, left ventricular angiography is useful for estimating left ventricular volume and ejection fraction. The cause of the chronic mitral regurgitation in individual patients influences the prognosis. In [B. Shipton and W. Wahba (2001)], a complete description of valvular heart diseases are discussed by [P.Stein (1981)].

In [Brian D. Hoit and Michael D. Faulx (2002)], several reparative surgical techniques with acceptable morbidity and mortality rates now exist. Although two-thirds of patients with rheumatic mitral valve disease have pathologic evidence of tricuspid valve involvement, clinically significant tricuspid disease, which generally affects young and middle-aged women, is much less common. Right atrial myxomas and sarcomas and obstructing metastatic tumors may produce hemodynamic changes that are indistinguishable from tricuspid stenosis. The neck veins are distended, and the earlobes may pulsate. Because venous distention may obscure the jugular pulse contour, it is important to elevate the patient's head. Contrast or color-flow Doppler echocardiography can visualize the tricuspid regurgitant jet. Although right ventriculography requires a catheter across the tricuspid valve, there is no significant contrast leak into the right atrium under normal circumstances. Patients with congenital pulmonic stenosis can expect a life expectancy comparable to that of the general population.

Author [H. Fletcher (1940)], once the diagnosis of hypertension is made and therapy instituted, elevated blood pressure can be lowered, reducing the risk of cardiovascular disease in the vast majority of patients. High blood pressure is uncommon under the age of 20; if present, it is usually associated with renal insufficiency, renal artery stenosis, or contraction of the aorta. The blood pressure should be measured, with a cuff of the appropriate size, after at least 5 minutes of rest in a seated or supine position. A complete evaluation of all prescription and nonprescription medications the patient is taking should be done to exclude any possible contribution to the elevation or any interaction that might limit a given drug's antihypertensive effects. The prognosis is significantly worse when LVH is present with any amount of blood pressure elevation. The use of all “stimulant” type weight reduction therapies should be strictly avoided, as they tend to elevate blood pressure. The fat substitutes or avoidance therapies do not raise blood pressure but have their own side effects.

Author [D.D. Greenwood, (1961)] presents new results in human identity verification via frequency analysis of cardiac sounds. More specifically, the focus is on pattern recognition approach based on a feature set of 13 Mel Frequency Cepstral Coefficients (MFCCs) extracted from the first (S1) and second (S2) heart sounds and a metric based on the power ratio of S1 to S2. The performance of an identification system based on frequency analysis of heart sounds depends both on the segmentation of S1 and S2 sounds and on the biometric features used in the subsequent matching phase.

Frequently, murmurs or alterations in the heart sounds are the only definitive signs of some types of heart disease, appearing long before stress on the cardiovascular system is sufficient to produce other signs and symptoms [J. Zwislocki, (1965)]. The PCG signals of five healthy, normal subjects and 20 patients with valvular and other cardiovascular defects were taken up for study using the techniques described.

In [Wang & Huang (1997)], the effects of each of the psychoacoustic facts are investigated to obtain their relative importance on a speaker-dependent automatic speaker recognition (ASR) system. The speech analysis attempts to model the speech auditory spectrum by an all-pole function. The auditory spectrum is generated from critical-band filtering, followed by equal loudness weighting and intensity loudness conversion. The recognizer is based on fixed-end point dynamic time warping algorithm. Author concluded that introducing perceptual weightings into short-time spectral estimates of speech signal reduce significantly the order of the ASR feature vector.

In [Su, Hu and Zhang (1994)], the authors discussed about the use of time-frequency methods in the detection heart murmurs and their analysis in phonocardiographic (PCG) signals. There are many potential causes of murmurs, a majority of which are non-pathological, and hence require no treatment. Heart sounds are non-stationary signals and short-time Fourier techniques are more appropriate for the non-stationary signals. The general processing strategy they adopt consists of initially segmenting the signal into individual frames. In [Su, Hu and Zhang (1994)], they found that the time-frequency methods are capable of detecting heart murmurs and also of yielding information vital to the classification of such murmurs. In [E. Zwicker and H. Fast (1990)], the authors present an analytical perspective on cardiac auscultation. They focus on disorders such as valvular disease and stenos is that are typically detected using auscultation rather than on disorders that are typically detected using ECG. An implicit assumption of their work, which is used in segmenting the signal, is that subjects have hearts that are normal from an electrophysiological perspective.

In [E. Zwicker and H. Fast (1991)], the authors have concluded that the MAAS Framework provides a set of tools for automatically segmenting the audio signal into intervals corresponding to important sections of the cardiac cycle.

2.2 PSYCHOACOUSTIC MODELING AND CONTENT BASED MULTI MEDIA INFORMATION RETRIEVAL

The field of psychoacoustics [H. Fletcher (1940)] [D.D. Greenwood (1961)][J. Zwislocki (1965)] [E. Zwicker and H. Fastl (1990)][E. Zwicker and U. Zwicker (1991)] has made significant progress toward characterizing human auditory perception and particularly the time-frequency analysis capabilities of the inner ear. Although applying perceptual rules to signal coding is not a new idea [J. T. Foote (1997)], most current audio coders achieve compression by exploiting the fact that "irrelevant" signal information is not detectable by even a well trained or sensitive listener.

In a physical and physiological basis for the interpretation of cardiac auscultation is performed by author [J. Zwislocki 1965]. Evaluation is based on primarily on the second sound and ejection murmurs.

In [B.R Glasberg and B.C Moore (2002)], a loudness model for steady sounds is described having the following stages: 1) A fixed filter representing transfer through the outer ear; 2) A fixed filter representing transfer through the middle ear; 3) Calculation of an excitation pattern from the physical spectrum; 4) Transformation of the excitation pattern to a specific loudness pattern; 5) Determination of the area under the specific loudness pattern, which gives overall loudness for a given ear; and 6) Summation of loudness across ears.

The model differs from earlier models in the following areas: 1) The assumed transfer function for the outer and middle ear; 2) The way that excitation patterns are calculated; 3) The way that specific loudness is related to excitation for sounds in quiet and in noise; and 4) The way that binaural loudness is calculated from monaural loudness.

The model is based on the assumption that sounds at absolute threshold have a small but finite loudness. This loudness is constant regardless of frequency and spectral content. It is also assumed that a sound at masked threshold has the same loudness as a sound at absolute threshold. The model accounts well for recent measures of equal-loudness contours, which differ from earlier measures because of improved control over bias effects. The model correctly predicts the relation between monaural and binaural threshold and loudness. It also correctly accounts for the threshold and loudness of complex sounds as a function of bandwidth [E. Zwicker and H. Fastl, (1990)]. A new model which can predict the loudness of brief sounds as a function of duration and the overall loudness of sounds that are amplitude modulated at various rates [B.R Glasberg and B.C Moore (2002)].

Concepts of psychoacoustics are the study and discussed in [J. P. Christopher (2005)]. It is defined as the subjective human perception of sounds. Alternatively it can be described as the study of psychological correlates of the physical parameters of acoustics. Classical reference [B.C.J More (1997)] which represents a set of algorithms for calculating auditory sensations including loudness, sharpness, roughness, softness, sound strength and intensity, and fluctuation strength. In this thesis work, it is extended for the heart sounds and murmurs. The classification of murmurs is characterized by using the psychoacoustic features and derives mathematical equations. In [S.S Stevens, J. Volkman and E. Newman (1937)], the authors present for the calculating the sharpness of tones in musical tones is modeled and used in the music analysis system. A survey on content-based music information retrieval: current directions and future challenges are studied and presented in [Kemp (1995), Remco, Tanase and Daniele (1999)].

In [Sebe, N., Lew, M.S., Zhou, X., and Huang, T.S. (2003)] state of art work [Lew (2001); Sebe, et al. (2003), J. T. Foote (1997)] done in content-based multimedia retrieval in the recent years are discussed thoroughly. An overview of trends and developments in the area of Content-Based Audio Indexing and Retrieval (CBAIR), during the past few years along with limitations and constraints of existing Query by Example (QBE) and Query By Humming (QBH) CBAIR systems is discussed. Different methods to represent musical objects, like feature-based representation, musical parameter-based representation; similarly based retrieval strategies, like feature based retrieval as well as melody or theme based retrieval of musical objects, in this paper. Moreover, some important issues regarding indexing and retrieval performance i.e. efficient indexing and retrieval complexity, in this area are discussed thoroughly in [J. T. Foote (1997)].

Audio parameters based systems have been extensively used for speak recognition and speaker identification systems for more than two decades and these systems are still popular in this area, but unfortunately, for CBAIR systems based especially for music retrieval systems; audio parameter based systems could not gain same popularity [Hafiz Malik, (1997)].

Author [J. Foote (1999)] presented an idea for the representation of an audio object by a template that characterizes the object in his purposed system. For construction of a template; an audio signal is first divided into overlapping frames of constant length then using signal processing techniques, for each frame a13-dimensional feature vector is extracted (12 Mel-Frequency Cepstral Coefficients plus Energy) at a 500Hz, and then these feature vectors are used to generate templates using tree-based vector quantizer trained to maximize mutual information (MMI). For retrieval, query is first converted in to template in the same way described earlier then for its similarity search template matching is applied which uses distance measure, and finally a ranked list is generated based on minimum distance. In this system performance of the system with Euclidean distance as well as Cosine distance, is also compared, and experimental results show that cosine distance performs slightly batter than Euclidean distance. This system may fail for music retrieval if either query is corrupted with noise or bad quality recorded.

A commercial company — Muscle fish group [Muscle Fish LLC. http://www.musclefish.com], in this system an audio object is characterized by its frame level and global acoustical and perceptual parameters. These features are extracted at frame level using signal processing techniques and globally using statistical analysis based on frame level features and musical features (for music signals only) using musical analysis. Frame level features consist of loudness, pitch, tone (brightness and bandwidth), MFCCs and derivative. Global features are determined by applying statistical modeling techniques on the frame level features that is, using Gaussian and Histogram Modeling techniques to analyze audio objects. For musical objects, musical features (i.e. rhythm, events and distance (interval)) are extracted using simple signal processing techniques like pitch tracking, voiced and unvoiced segmentation and note formation.

For indexing, multidimensional features space is used. For retrieval, distance measure is used and to improve the performance, a modified version of query-point-expansion [B. Scharf (1970) and R. Hellman (1972)] technique is used, but here expansion for the refinement of the concept if achieved by standard deviation and mean of the objects in the expected region and works for QBE only.

Authors [Khokhar and Li (2000), Keisler and Wold (1995), Joo (1982)] represented an audio object by 3-dimensional feature vector for their system. Feature vectors are extracted at frame level using Discrete Wavelet Transform (DWT). They applied 4-level DWT decomposition to audio signal, then in transformed domain variance, zero-crossing rate and mean of wavelet coefficients are determined to form feature vectors. For indexing structure, B-tree structure [Guttmann, (1984), Comer (1979)] is used, which is constructed using clustering technique along with multiresolution property of the wavelet transform. Similarity search is applied using weighted Euclidean distance, and based on minimum distance a ranked target list is retrieved for the desired query.

Authors [R. Subramanya and A. Youseef (1998)] presented a signal processing based approach using Discrete Wavelet Transform (DWT) for feature vector of an audio object. First of all an audio signal is decomposed using 4-level DWT and then form wavelet coefficients, feature vector is formed using all approximate coefficients and 20%-30% of detail coefficients of all levels obtained during wavelet decomposition. For query processing same procedure is applied. It is not clear about indexing technique but used Euclidean distance measure for similarity.

At this point it is important to note that the feature based similarity search engines were useful in a variety of contexts [Remco et al. (1999), Wang et al., (1998)] such as searching video shots with similar visual content and motion or for DJs searching for music with similar rhythms [Foote (1999]. Intuitively, the most pertinent applications are those where the basic features such as color and texture in images and video; or dominant rhythm, melody, or frequency spectrum in audio [Foote (1999)] are highly correlated to the search goals of the particular application.

2.3 POSITIONING OF THESIS

The survey of audio information retrieval (AIR) work clearly shows that, the content-based AIR techniques and algorithms have been successfully used for the music and song analysis, song classification, music genre classification and music retrieval applications. The spectral features and content-based algorithms and DSP techniques are extensively for the searching and retrieval of audio objects and related applications.

The previous AIR research motivated the current research work to extend the content based AIR algorithms, DSP approaches based on feature vectors (spectral and perceptional) for the searching and retrieval of heart sounds and murmurs. In particular, the proposed research work explores audio content-based algorithms and their applicability to for the searching, retrieval and classification of heart sound and murmurs. The performance of the algorithms in terms of relevance, recall and precisions is explored and contributes to the audio content based algorithms for heart sounds and murmurs.

In general, during auscultation, the doctors listen to the heart sounds and murmurs and classification is mainly based on psychoacoustic perception. The research psychoacoustics and audio perception motivated the thesis work and developed a psychoacoustic modeling and retrieval of heart sound and murmurs based on the psychoacoustic principles. The content-based retrieval and classification algorithms are explored and propose a set of classification and retrieval algorithms based on the audio-content and audio similarity measures that provide a qualitative and quantitative reasoning framework for the interpretation of heart diseases. The distance measures used in the pattern classification (e.g., Euclidian distance) are also useful for the similarity measures and introduce perception vector based similarity measures for the classification and retrieval applications. This thesis explores novel techniques and distance measures called – psychoacoustic distance measure used for the classification, indexing, and searching and retrieval application of heart sounds. These algorithms and techniques are implemented and experimentally validated with reference heart and murmur databases.

CHAPTER – 3
HEART SOUNDS AND MURMURS

3.0 PRELIMINARIES ON HEART SOUNDS AND MURMURS

The stethoscope is a recognized icon for the medical profession, and for a long time, physicians have relied on auscultation for detection and characterization of cardiac diseases. New advances in cardiac imaging have however changed this picture. Echocardiography and Magnetic Resonance Imaging (MRI) have become so dominating in cardiac assessment that the main use of cardiac auscultation is nowadays a preliminary test in the primary health care. Basically, all patients presenting anything but normal auscultatory findings are sent to a cardiology clinic for further investigations. Decision support systems based on heart sounds and murmurs would improve the accuracy of auscultation by providing objective additional clinical information. The aim of this thesis is to develop signal processing algorithms able to extract such information, use them as features for the, searching, retrieval and classification of heart sounds and murmurs.

Aristotle found the heart to be the seat of intelligence, motion and sensation. Other organs surrounding the heart, such as the brain and the lungs, merely existed as cooling devices. Since the fourth century BC, our understanding of the heart has changed from being an all-embracing organ to a highly specialized device whose purpose is to propel blood. Knowledge about auscultation has evolved alongside with discoveries about heart function. Robert Hooke (1635–1703), an English polymath, was the first to realize the diagnostic potential of cardiac auscultation. When Ren´e Laennec (1781–1826) invented the stethoscope in 1816, cardiac auscultation became a fundamental clinical tool [Michael. T (1997)] and remains so today. Normally there are two heart sounds, S1 and S2, produced concurrently with the closure of the atrioventricular valves and the semilunar valves, respectively. A third and a fourth heart sound, S3 and S4, might also exist. Additionally, a variety of other sounds such as heart murmurs or adventitious sounds may be present. Heart murmurs can be innocent or pathologic, and they are especially common among children (50 - 80%) of the population has murmurs during childhood, but only about 1% of these murmurs are pathological [A. N. Pelech (2004)]) and in the elderly prevalent estimates range from 29 – 60% [D. K. Dey, V. Sundh, and B. Steen, (2004)]. Most common are murmurs originating from the left side of the heart, especially aortic stenosis (AS) and mitral insufficiency (MI).

Auscultation is highly subjective and even the nomenclature used to describe the sounds varies amongst doctors. Unfortunately, the auscultatory skills amongst doctors and physicians demonstrate a negative trend. During auscultation murmurs can be detected. Performing auscultation is however difficult since it is based on the physician's ability to perceive and interpret a variety of low-intensity and low-frequency sounds. A plot of sound pressure (Pa - Pascal) against the frequency (Hz) is shown in Figure 3.1 and used as a reference in audio perception studies, including speech perception. Typical speech perception range falls in the range of 256 – 2048 Hz and 1 Pa to 10^{-2} Pa. The human speech is in the human perception range and does not pose serious perception challenges. However, the heart sound and murmurs fall in the range of 8 – 500 Hz with a pressure range of 10^{-2} (Pa) to 10^{-5} (Pa) and falls at the threshold of human hearing. Therefore, these heart sounds and murmurs cannot be heard using acoustical or electronic stethoscope as human beings cannot hear in the range of low-intensity and low-frequency sounds. The best way is to plot the phonocardiographic signal as function of time or frequency and enable it for visual inspection so that the heart sounds and murmurs can be visualized avoiding the limitation of the human hearing.

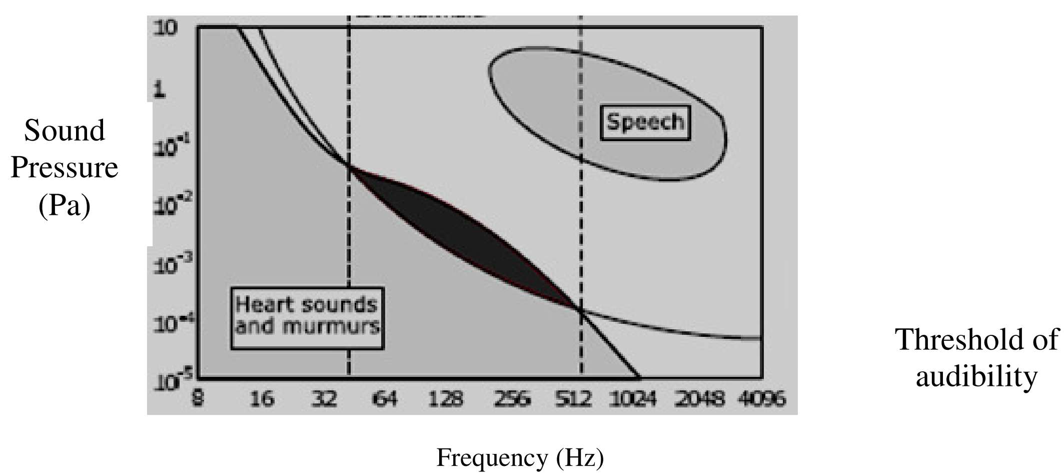


Figure 3.1 Relationship between the acoustic range of cardiac sounds and the threshold of audibility of the human ear adopted from [Leatham (1957)].

3.1 CARDIOVASCULAR ANATOMY AND PHYSIOLOGY

Heart sounds and murmurs arise as a consequence of turbulent blood flow and vibrating cardiovascular structures. In this section, the principles of anatomy and physiology that are necessary to understand how the cardiac sounds are related to physiological events is reviewed. It also focuses on the review of the electrical and mechanical operation of the healthy heart, along with the most important interactions within the cardiovascular system. The coupling between the cardiac system, the vascular system and the respiratory system is very interesting since it renders continuous, non-invasive and non-intrusive monitoring of respiration and blood pressure changes possible. The most important parameters governing mechanical activity are blood pressure, tension in the heart or in adjacent vessels, ventricular volume, blood flow velocity and movement as well as deformation of the heart wall [H. Vermarien and J. G. Webster (2006)]. Many of these parameters can only be measured with sophisticated equipment. However, since the mechanical events cause vibrations that are propagated to the chest surface, information about the working status of the heart can be obtained by auscultation.

There are basically two types of sounds originating from the heart; heart sounds and murmurs. A preliminary example showing a recorded PCG signal, containing the two normal heart sounds S1 and S2, is illustrated in Figure 3.2 along with an ECG. Murmurs can be of both pathological and physiological origin and arise as a consequence of increased blood flow in the heart. High velocities can be completely normal, especially amongst children, but it may also be due to a pathological narrowing in the blood's pathway. A common cause of such obstructions is valvular heart diseases and path physiology of the most common valvular dysfunctions is necessary for the characterization of heart diseases.

The cardiovascular system is designed to establish and maintain a mean systemic arterial pressure sufficient to transport nutrients, oxygen and waste products to and from the cells, while preserving regulatory flexibility, minimizing cardiac work and stabilizing body temperature and pH to maintain homeostasis [J. Katz et el.,(1992)]. The main components of the cardiovascular system are the heart, the blood, and the blood vessels.

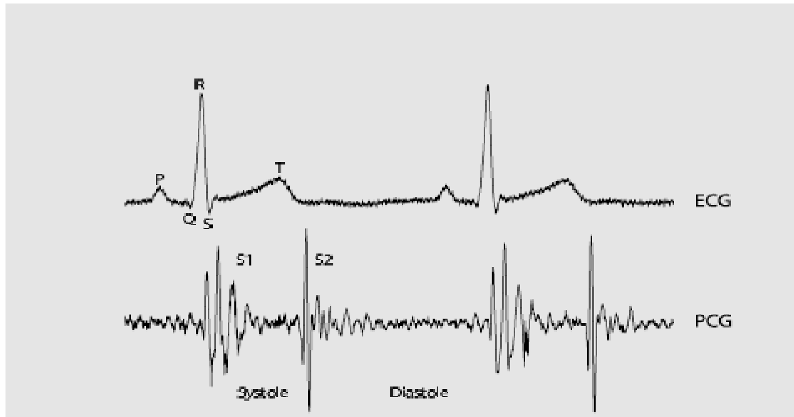


Figure 3.2 An electrocardiogram (ECG) and a phonocardiographic (PCG) signal from a healthy person without murmurs [R.J. Rangayyan, R.J.Lehner (1988)].

The primary task of the heart is to serve as a pump propelling blood around the circulatory system. When the heart contracts, blood is forced through the valves, at first from the atria to the ventricles and then from the ventricles out through the body as shown in Figure 3.3. There are four heart chambers, the right and left atria and the right and left ventricles. From a simplistic point of view, the two atria mainly act as collecting reservoirs for blood returning to the heart while the two ventricles act as pumps ejecting blood out of the heart through the body. The pumping action of the heart is divided into two phases, systole when the ventricles contract and ejects blood from the heart, and diastole, when the ventricles are relaxed and the heart is filled with blood. Four valves prevent the blood from flowing backwards. The atrioventricular valves (the mitral and tricuspid valve) prevent blood from flowing back from the ventricles to the atria and the semilunar valves (aortic and pulmonary valves) prevent blood from flowing back towards the ventricles once being pumped into the aorta and the pulmonary artery, respectively. Deoxygenated blood from the body enters the right atrium, passes into the right ventricle and is ejected out through the pulmonary artery on its way to the lungs. Oxygenated blood from the lungs re-enter the heart in the left atrium, passes into the left ventricle and is then ejected out through the body.

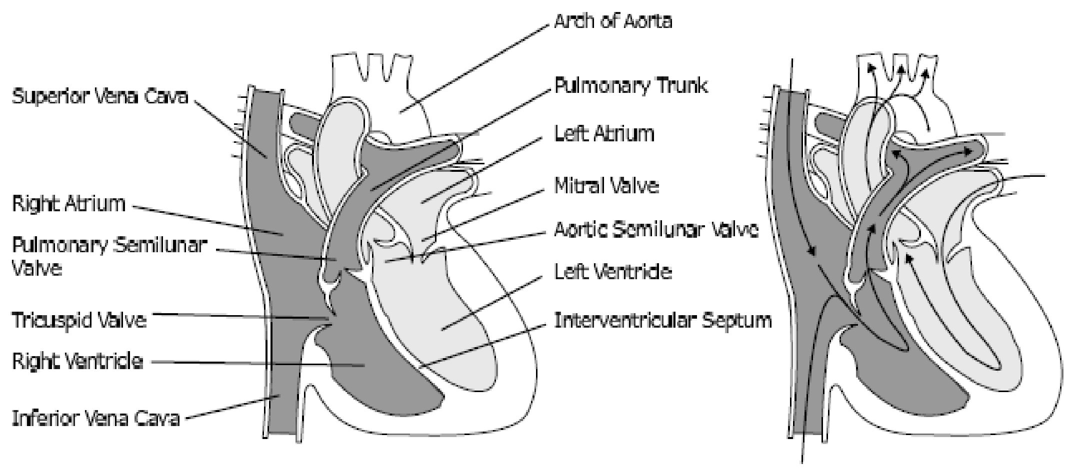


Figure.3.3 Anatomy of the heart (left figure) and the blood flow pathways through left and right side of the heart (right figure).

3.1.1 THE HEART VALVES

The atria are separated from the ventricles by the fibrous skeleton of the heart [J. Katz et al., (1992)]. There is one fibrous ring around each of the four valves, but the rings are fused together into a single fibrous framework. The skeleton has several physiological functions. It provides a foundation to which the valves and the arteries attach, it prevents overstretching of the valves as blood passes through them and it electrically isolates the atria from the ventricles. All four heart valves have flaps, called leaflets or cusps, which open to let the blood flow through and close to prevent it from flowing backwards. The valves and their leaflets are illustrated in Figure 3.4. The mitral and tricuspid valve leaflets are connected via the chordae tendineae and papillary muscles to the ventricular wall. The papillary muscles contract at the same time as the ventricles contract, thus pulling the chordae tendineae downwards and preventing the valve leaflets into the atria. The semilunar valves both have three cusps consisting of connective tissue reinforced by fibers. These valves do not have chordae tendineae, instead the shape of the cusps prevent any form of prolapse.

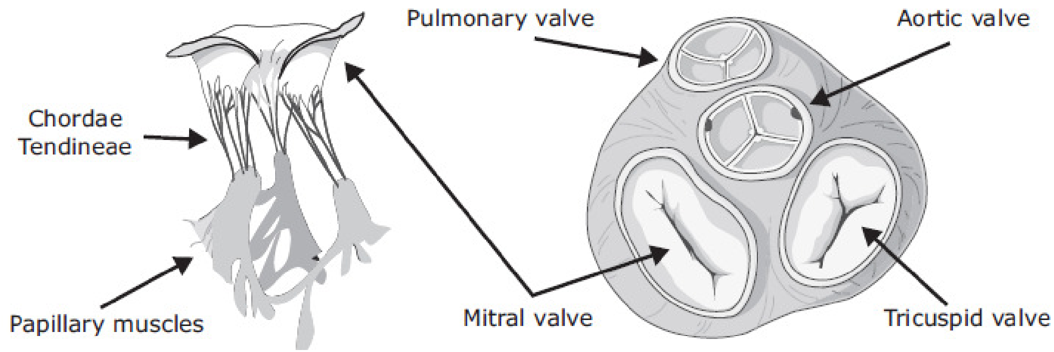


Figure 3.4 Illustration of the mitral valve and its associated chordae tendineae and papillary muscles (left) and the heart valves and the fibrous rings surrounding each valve (right).

3.1.2 THE CARDIAC ELECTRICAL SYSTEM

Cardiac muscle cells can possess at least four properties, automaticity (the ability to initiate an electrical impulse), conductivity (the ability to conduct electrical impulses), contractility (the ability to shorten and do work) and lusitropy (the ability to relax) [R. M. Berne and M. N. Levy (1997)]. Cells in different areas of the heart are specialized to perform different tasks; all cells possess the conductivity property, the working cells are mainly able to contract and relax while the cells governing the electric systems are adapted to automaticity and conductivity. The pumping action of the heart is synchronized by pacemaker cells, concentrated in the sinoatrial node (located in the right atrium), the atrioventricular node (located in the wall between the atria) and in the His-Purkinje system (starting in the atrioventricular node and spreading over the ventricles), see Figure 3.5. An action potential generated in the sinoatrial node (which normally controls the heart rate) will spread through the atria and initiate atrial contraction. The atria are electrically isolated from the ventricles, connected only via the atrioventricular node which briefly delays the signal. The delay in the transmission allows the atria to empty before the ventricles contract. The distal part of the atrioventricular node is referred to as the Bundle of His. The Bundle of His splits into two branches, the left bundle branch and the right bundle branch, activating the left and the right ventricle, respectively. The action potential spreads very quickly through the ventricle due to the fast His-Purkinje cells, causing almost immediate synchronous excitation of the entire ventricular wall [Pelech, (2004)].

3.1.3 THE ELECTROCRADIOGRAM (ECG)

Cardiac action potentials are conducted to the body surface, where they can be measured as an electrical potential that varies with the current flow through the heart. Action potentials associated with different cardiac regions are illustrated in Figure 3.5, along with a typical ECG waveform measured from the body surface.

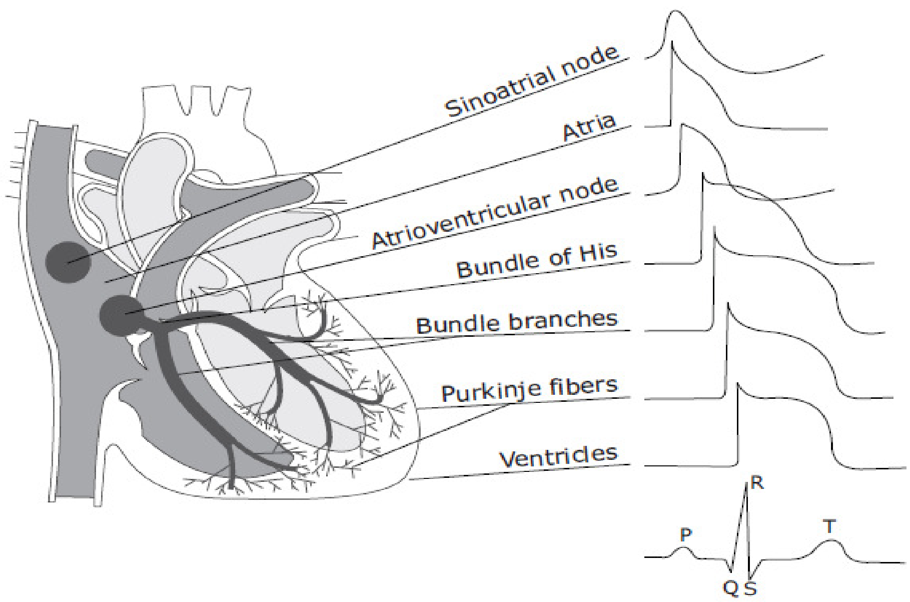


Figure 3.5 Morphology and timing of action potentials from different regions of the heart are illustrated in the right-hand side of the figure [L. S¨ornmo and P. Laguna, (2005)].

The ECG can be seen as a projection of a dominant vector (represented by the summation in time and space of the action potentials from each muscle cell) onto a lead vector, whose direction is defined by the position of the measurement electrodes in relation to the heart [L. S¨ornmo and P. Laguna (2005)]. The ECG describes the different electrical phases of the heart, where depolarization of the atria gives rise to the P-wave, depolarization of the ventricles combined with repolarization of the atria results in the QRS-complex and repolarization of the ventricles results in the T-wave.

3.1.4 THE CARDIAC CYCLE AND THE PRESSURE-VOLUME

The blood pressure within a chamber increases as the heart contracts, generating a flow from higher pressure areas towards lower pressure areas. The work diagram of the heart, illustrated in Figure 3.6 for the left ventricle, is referred to as a pressure volume (PV) loop [J. Katz (1992]. The following discussion applies to the left side of the heart, but the key concepts are similar for the right side.

When left atrial pressure exceeds the pressure in the left ventricle, the mitral valve opens (A) and the atrium empties into the ventricle (filling). During the rapid filling phase, venous blood from the lungs enters the atrium, and as the pressure gradient between the atrium and the ventricle levels out (reduced filling phase), a final volume of blood is forced into the ventricle by atrial contraction. When tension develops in the ventricular wall, increased intraventricular pressure will force the mitral valve to shut (B). The pressure stretching the ventricle at this moment is called preload. The amount of pressure exerted is determined by the duration of ventricular diastole together with the venous pressure. Within limits, the more the heart is stretched during diastole, the more vigorous the contraction will be in systole. Since, the heart is contracting while all valves are closed, ventricular pressure will increase whereas the volume remains unchanged (isovolumic contraction). The first heart sound originates from events related to the closing of the mitral valve (B) and the opening of the aortic valve (C). The ventricular pressure required to open the aortic valve is called “after load”, a parameter which, consequently, is affected by arterial blood pressure.

As blood is ejected from the heart, ventricular pressure decreases, and when it falls below the aortic pressure, the aortic valve closes again (D). In association with valve closure, S2 is heard. The end-systolic pressure-volume ratio is a clinical measure of cardiac muscle performance referred to as myocardial contractility. Again all valves are closed, but this time the pressure will decrease while the volume remains unchanged. This phase, called isovolumetric relaxation, will complete the loop and start a new heart cycle.

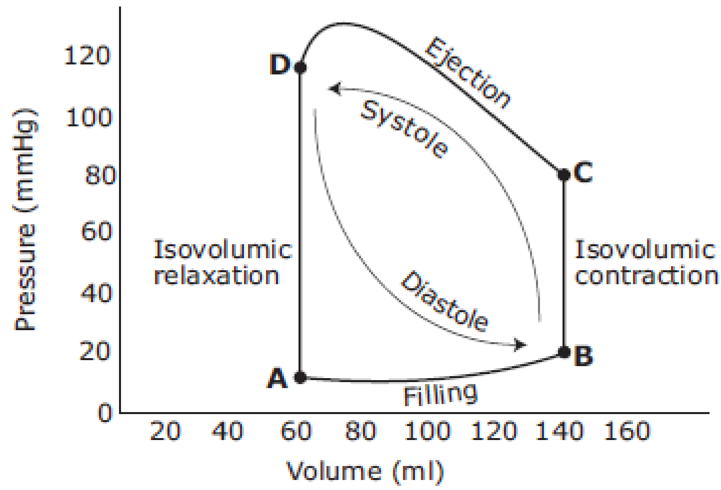


Figure 3.6 Work diagram (pressure-volume loop) of the left ventricle.

The PV-loop illustrates the changing pressures and flows within the heart; however, it has no time scale. In Figure 3.8, Wiggers diagram demonstrates the temporal correlations between electrical and acoustical events in the left side of the heart over one cardiac cycle [J. Katz et el., (1992)]. The electrical R-wave, representing ventricular depolarization, precedes the beginning of ventricular contraction. The ventricular contraction causes a rapid rise in the left ventricular pressure. As soon as the ventricular pressure exceeds the atrial pressure, the mitral valve closes (B in the PV-loop). This is when S1 is heard. When the ventricular pressure exceeds the aortic pressure, the aortic valve opens (C in the PV-loop), and the blood flows from the ventricle to the aorta. At the end of blood ejection, the pressure in the ventricle falls below the aortic pressure, and the aortic valve closes (D in the PV-loop), giving rise to S2. The pressure in the ventricle drops steeply, and when it falls below the atrial pressure, the mitral valve opens (A in the PV-loop), and the rapid filling phase begins. The rapid filling phase might cause an impact sound, the third heart sound (S3), when blood collides with the ventricular wall. Similarly, atrial systole may also produce an audible forth heart sound (S4).

3.1.5 VALVULAR HEART DISEASES

Valvular heart diseases are more common in the mitral and aortic valves since the left side of the heart sustains higher pressures and greater workloads. There are two major problems that may compromise the functionality of the valves, stenosis and insufficiency [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005)]. In stenosis the leaflets become rigid, thickened or fused together, reducing the opening through which the blood passes from one chamber to another. The obstructed flow gives rise to an accumulation of blood in the chamber, forcing the heart to work harder in order to pump the blood. In insufficiency (or regurgitation) the valves fail to close properly why a portion of the ejected blood flows backward. For example, if the mitral valve is unable to close properly, some of the blood will leak back into the left atrium during systole. Valvular stenosis and insufficiency gradually wear out the heart. At first, the heart muscle thickens (hypertrophy) and the heart enlarges (dilatation), thus compensating for the extra workload and allowing the heart to supply an adequate amount of blood to the body. Over time, the overdeveloped heart muscle may lead to a functional degradation and heart failure.

Aortic stenosis (AS) is an obstruction between the left ventricle and the aorta. The obstruction may be in the valve (valvular), above the valve (supravalvular) or below the valve (subvalvular). The most common causes are congenital abnormality, rheumatic fever, or calcific degeneration or deposits of calcium on the valve. A typical audio recording of aortic stenosis is shown Figure 3.7

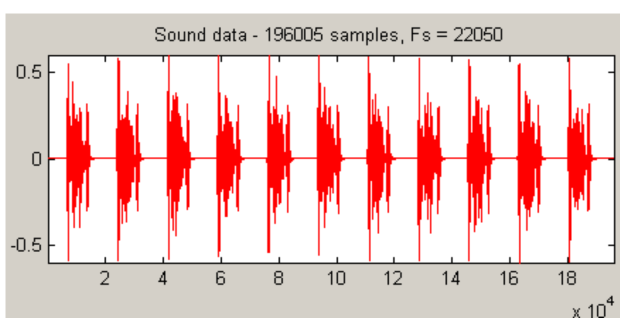


Figure 3.7 A phonocardiogram of aortic stenosis.

In the presence of an obstruction, a pressure gradient develops between the left ventricle and the ascending aorta. As a response to the increased left ventricular pressure, hypertrophy is developed. Since left ventricular hypertrophy offers increased resistance to filling, preload is elevated (through strong atrial contractions). Eventually, the increased left atrial pressure produces pulmonary edema, leading to increased pressures in the right side of the heart, increased systemic venous pressure and peripheral edema [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005)].

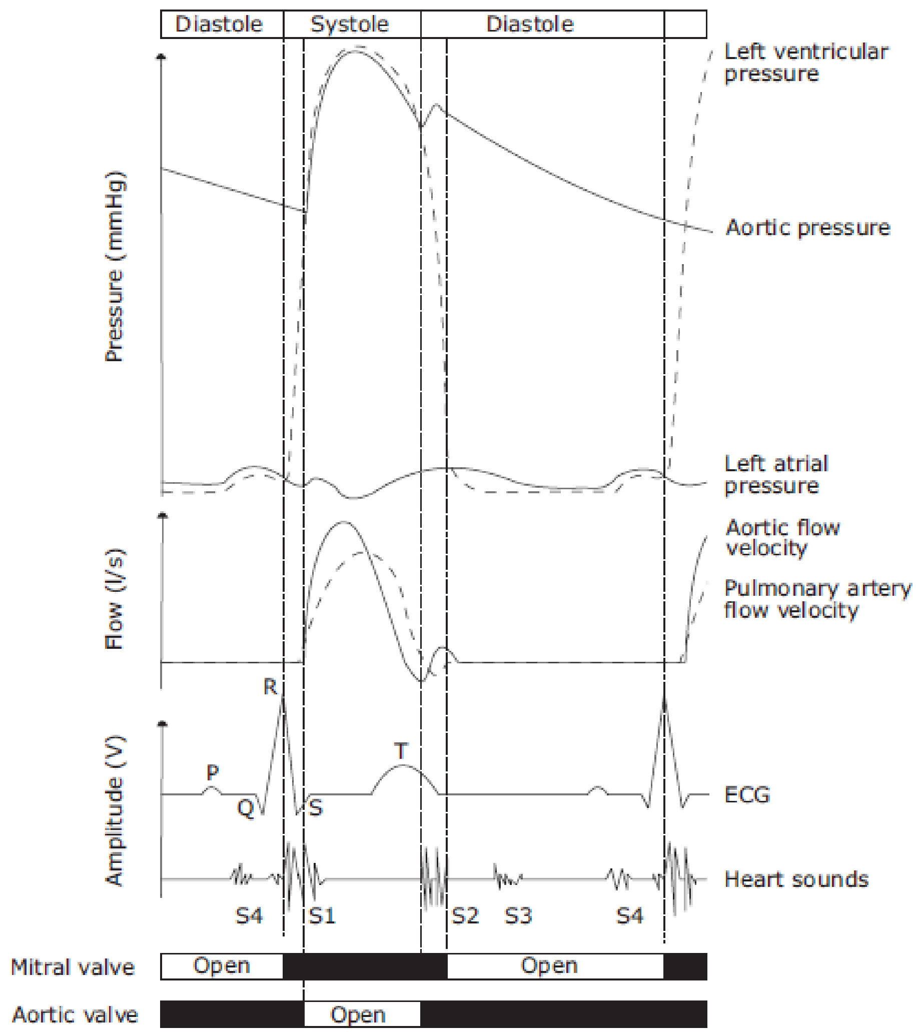


Figure 3.8 Wiggers diagram, showing pressures and flows in the left side of the heart over one heart cycle and how they relate to electrical (ECG) and acoustical (PCG) activity.

Aortic insufficiency (AI) refers to an incompetent aortic valve allowing blood to flow back into the left ventricle during diastole when the ejection is complete. In its acute form, aortic regurgitation usually occurs as a result of infective endocarditis that destroys the valve's leaflets. The chronic form, which is more common, is usually a consequence of widening of the aorta in the region where it connects to the valve. In either case, the constant leaking of blood results in increased left ventricular diastolic pressure, increased left atrial pressure and eventually heart failure and pulmonary edema [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005)].

Mitral stenosis (MS) is a narrowing or blockage of the mitral valve, often as a result of rheumatic fever. The narrowed valve causes blood to back up in the left atrium instead of flowing into the left ventricle and results in an increase in the pressure in the left atrium. This pressure is transmitted back through the pulmonary veins, causing pulmonary edema and consequent problems in the right side of the heart [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky Hurst's (2005)]. A typical phonocardiogram of mitral stenosis is shown in Figure 3.9.

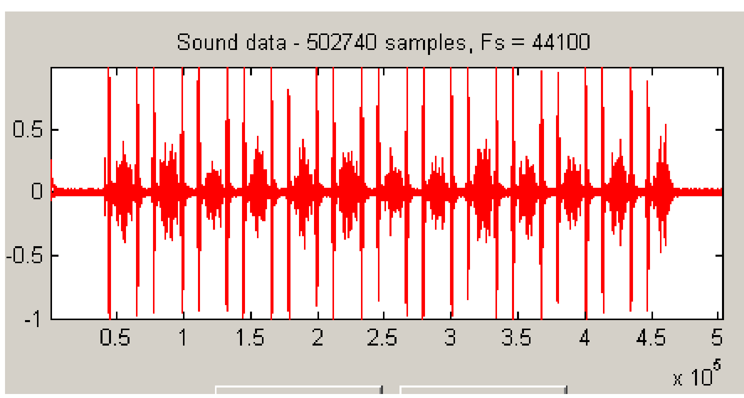


Figure 3.9 A phonocardiogram of mitral stenosis.

Mitral insufficiency (MI) is an abnormal leaking of blood from the left ventricle into the left atrium of the heart. The most common causes are myxomatous degeneration of the valve, annulus dilatation, dysfunction of the papillary muscles or rupture of the chordae tendineae. The amount of blood that flows back into the atrium is called a regurgitant volume. The regurgitant volume depends on three factors: the area of the leaking orifice, the pressure gradient between the chambers and the regurgitant duration. Since blood is ejected into the left atrium instead of out through the aorta, the forward stroke volume decreases. In response, the heart compensates by increasing the total stroke volume and the heart rate, and by eccentric hypertrophy. A phonocardiogram of mitral insufficiency is displayed in the Figure 3.10.

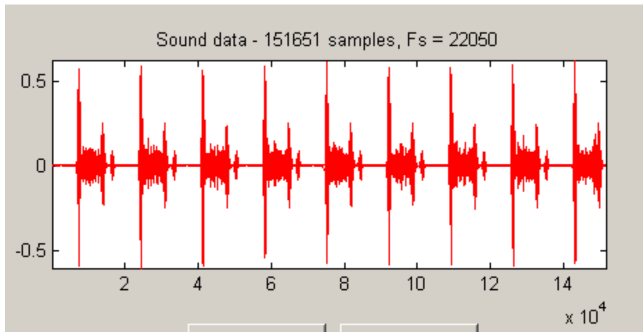


Figure 3.10 A phonocardiographic recording of mitral insufficiency.

The atrium will increase its force of contraction in order to maintain ventricular filling. The consequent increase in atrial pressure may lead to pulmonary congestion and edema. Tricuspid and Pulmonic stenosis and regurgitation only account for a small amount of the valve diseases and is most often secondary to disease in the left side of the heart. Abnormalities of the tricuspid valve are generally caused by rheumatic fever [K. Dumont (2004)]. A reference phonocardiographic recording of pulmonic stenosis is shown in the Figure 3.11.

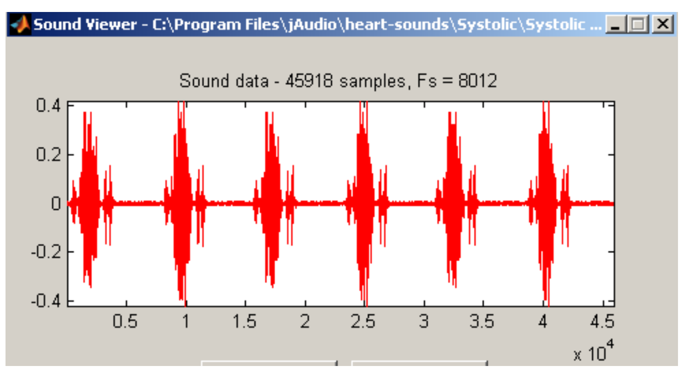


Figure 3.11 A phonocardiographic recording of pulmonic stenosis.

Rheumatic Fever (RF) is an inflammatory condition that often starts with strep throat or scarlet fever. Though the disease is rarely fatal during the acute stage, it may lead to rheumatic valvular disease, a chronic and progressive condition that causes cardiac disability or death many years after the initial event [Ara G. Tilkian and Mary Boudreau Conover (2001)]. The damage is not caused by the bacteria themselves, but by an autoimmune response - a process in which the body mistakenly begins to damage its own tissues.

Infective Endocarditis (IE) is a disease caused by microbial infection of the endothelial lining of the heart [Ara G. Tilkian and Mary Boudreau Conover (2001)]. The infection can cause vegetations on the heart valves, which sometimes conjures new or altered heart murmurs, particularly murmurs suggestive of valvular regurgitation [Ara G. Tilkian and Mary Boudreau Conover (2001)].

Myxomatous degeneration is a pathological weakening, mainly affecting the mitral valve. This dysfunction stems from a series of metabolic changes, causing the valve's tissue to lose its elasticity while becoming weak and covered by deposits.

Calcific degeneration is a hardening formed by deposits of calcium salts on the valve. This type of tissue degeneration usually causes aortic stenosis, a narrowing of the aortic valve [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005)].

Congenital anomalies are abnormal structures in the heart that occur by birth. The most common congenital valve defect is bicuspid aortic valves (two leaflets instead of three). Although not a valvular disease, Septal defects (an abnormal passage between the left and the right side of the heart) should also be mentioned since they are also congenital anomalies which gives rise to murmurs as shown in Figure 3.12.

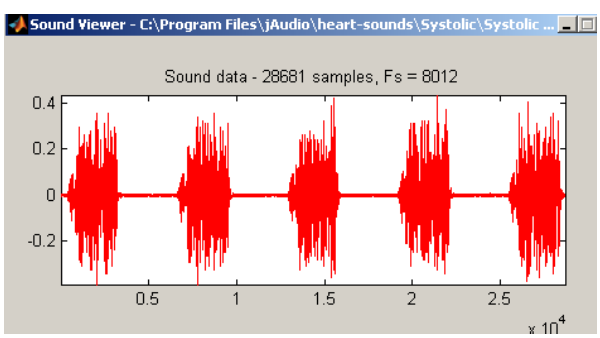


Figure 3.12 A phonocardiographic recording of septal defect.

Ventricular Septal Defect (VSD) is generally considered to be the most common type of malformation, accounting for 28% of all congenital heart defects [R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005)]. The other pathological causes for the murmurs and heart valve diseases are from the result from other heart diseases, particularly coronary artery disease or myocardial infarction. These conditions can cause injury to one of the papillary muscles that support the valves, or annulus dilatation, so that the valve does not close properly.

3.2 AUSCULTATION AND PHONOCARDIOGRAPHY

The technique of deciphering the sounds from the body based on their intensity, frequency, duration, number and quality is called auscultation [Y-T. Zhang, G. Chan, X-Y. Zhang and L. Yip. (2006)]. The acoustical signal is affected by a chain of transfer functions before the physician's actual decision-making process starts. The signal transmitted from the sound source is propagated through the human body, where the sound waves are both reflected and absorbed. The most compressible tissues such as lung tissue and fat contribute most to the absorption. Low frequencies are less attenuated compared to high frequencies, but the high frequencies are easier to perceive.

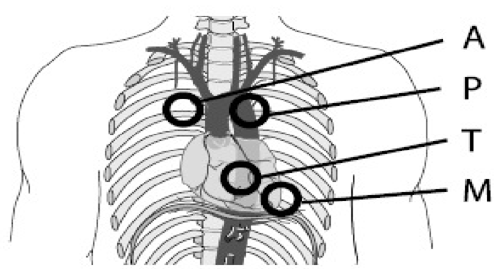


Figure 3.13 The traditional auscultatory areas on the chest (M refers to the mitral area, T the tricuspid area, P the pulmonic area, and A the aortic area).

The consequences of the attenuation are therefore hard to predict. To reduce the effect of thoracic damping, certain areas of cardiac auscultation have been defined. In these locations, as shown in Figure 3.13, the sound is transmitted through solid tissues or through a minimal thickness of lung tissue. The traditional areas of auscultation where the radiated sound intensity from each of the four heart valves is maximized and are described below defined [Ara G. Tilkian and Mary Boudreau Conover (2001)].

• **Mitral area:** The cardiac apex.

• **Tricuspid area:** The fourth and fifth intercostal space along the left sternal border.

• **Aortic area:** The second intercostal space along the right sternal border.

• **Pulmonic area:** The second intercostal space along the left sternal border.

3.2.1 TERMINOLOGY FOR DESCRIBING CARDIAC SOUNDS

Of the two normal heart sounds, S1 is louder, longer and lower pitched compared to S2. While S1 and S2 are referred to as *tones*, murmurs are characterized by a sound most easily described as "*noise-like*". During auscultation, murmurs are described by a number of factors: timing in the cardiac cycle, intensity on a scale of I-VI, shape, frequency, point of maximal intensity and radiation. A grade I murmur is very faint and heard only with special effort while grade VI is extremely loud and accompanied by a palpable thrill. When the intensity of systolic murmurs is crescendo-decrescendo shaped and ends before one or both of the components of S2, it is assumed to be an ejection murmur. Murmurs due to backward flow across the atrioventricular valves are of even intensity throughout systole and reach one or both components of S2.

If the regurgitant systolic murmur starts with S1 it is called holosystolic and if it begins in mid or late systole it is called a late systolic regurgitant murmur. Besides murmurs, ejection clicks might also be heard in systole. It is caused by abnormalities in the pulmonary or aortic valves. Different murmurs, snaps, knocks and plops can also be heard in diastole, but such diastolic sounds are difficult characterize and interpret these heart sounds and murmurs [Erickson (1997), Bell and Vecchione (1930)].

3.3 PHONOCARDIOGRAPHY

A graphical representation of the waveform of cardiac sounds is called a phonocardiogram, and the technique used to capture the sound signal is referred to as phonocardiography. For understanding, a typical recording of the PCG is shown in Figure

3.14. This technique allows a visual interpretation of the cardiac sounds, thus allowing thorough investigation of temporal dependencies between mechanical processes of the heart and the sounds produced. Today, PCG is mainly used for teaching and training purposes [H. Vermarien. G. Webster (2006)], but since new electronic stethoscopes make the recording procedure much easier, PCG might make a comeback in clinical practice.

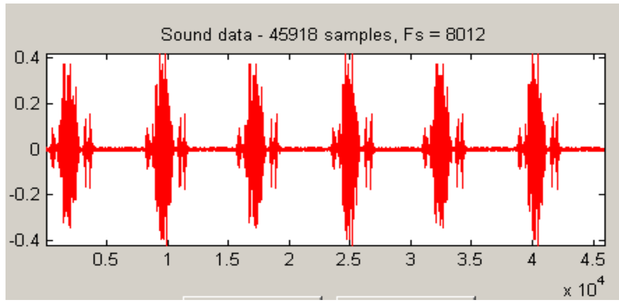


Figure 3.14 A PCG recoding of systolic pulmonary stenosis.

3.3.1 PHONOCARDIOGRAPHY TECHNIQUE

The auscultation of the heart gives the clinician valuable information about the functional integrity of the heart. Additional details can be gathered when the temporal relationships between the heart sounds and the electrical and mechanical events of the cardiac cycle are compared. This approach to the analysis of heart sounds using a study of the frequency spectra is known as phonocardiography. The phonocardiogram is a device capable of obtaining heart sounds and displaying the obtained signals in the form of a graph drawn with the signal amplitude in one axis and with time in the other.

The graph so obtained remains relatively obscure, but can be read by doctors, who are quite conversant with the characteristics of the graph [Pena C. et al., (1995)]. The electrocardiograph (ECG) signal is used as the reference wave in the analysis of heart sounds. The reference wave helps the observer in distinguishing the various parts of the heart cycle. The waveform generated by a heartbeat is, in comparison with a normal sine wave, a complex one, but the two can be related. In fact, one of the most powerful tricks in an engineer's mathematical tool bag is Fourier's theorem which maintains that any signal, no matter what it looks like, can be copied exactly by adding together many sine waves of different frequencies. It then follows that Fourier analysis can break down a complex signal into its sine wave components. If the signal has unwanted frequencies, these can then be filtered out. However, for heartbeat signals there is no way to tell which component is noise and which one a heartbeat. Hence, simple filtering doesn't work. This eventually led us to a spectral analysis of phonocardiography [Pena C. et al (1995)].

Cardiac sounds are the heart sounds that are short-lived bursts of vibrational energy having a transient character. They are primarily associated with valvular and/or ventricular vibrations. Both their site of origin and their original intensity governs the radiation of the heart sounds to the surface of the chest. There are four separate basic sounds that occur during the sequence of one complete cardiac cycle. Table 3.1 shows the principle characteristic of the heart murmurs derived spatially and have clinical importance in diagnosis in heart valve abnormalities and describes the different sound and the origin of each one.

Table 3.1 Spatial characteristics of diagnosing valve disease from heart murmurs.

Heart Sound	**Sound occurs during**	**Associated with**
S1	Isovolumetric contraction	Mitral and tricuspid valves closure
S2	Isovolumetric relaxation	Aortic and pulmonary valves closure
S3	Early ventricular filling	Sound Rapid ventricular filling in early diastole. Normal in children.
S4	Atrial contraction	Associated with stiff, low compliant ventricle.

Table 3.2 Murmurs and their psychoacoustic or perceptual features [Leathm, (1975)].

Heart Sound or Murmurs	Sound/Acoustic Properties
Aortic Stenosis (AS)	High pitch, high energy envelope,
Mitral Regurgitation (MR)	High pitch, high energy envelope with clear
Third Heart Sound (TS)	Faint heart sound after second heart sound.
Fourth Heart Sound (FS)	Faint heart sound after third heart sound.
Early Systolic Murmur (ESM)	Systolic, early cycle of S1, high pitch, high freq.
Late Systolic Murmur (LSM)	Systolic, late cycle of S1 or S2, high pitch, high freq.
Ejection Click (EC)	High energy pulse of 2-5 ms.
Diastolic Rumble (DR)	Rubbing sound
Atrial Septal Defect (ASD)	Gushing sound with high pitch
Patient Ductus Arteriosus (PDA)	Gushing sound with low pitch and inaudible
II Heart Sound Split (2SS)	Clear split sound of duration 2- 48 ms after s1.
III Heart Sound Split (3SS)	Clear split sound of duration 5 - 50 ms after s2.
Diastolic Summation Gallop (DSG)	Galloping sound
Diastolic Tricuspid Stenosis (STS)	High pitch and rhythmic
Diastolic Ventricular Gallop (SVG)	Galloping with high intensity & reducing with time.
Ejection Murmur (EM)	High intensity and high pitch ejection sound.

Cardiac murmurs are vibrations caused by turbulence in the blood as it flows through some narrow tube. A murmur is one of the more common abnormal phenomena that can be detected with a stethoscope - a somewhat prolonged "whoosh" that can be described as blowing, rumbling, soft, harsh, and so on. Murmurs are sounds related to the non-laminar flow of blood in the heart and the blood vessels. They are distinguished from basic heart sounds in that they are noisy and have a longer duration. While heart sounds have a low frequency range and lie mainly below 200 Hz, murmurs are composed of higher frequency components extending up to 1000 Hz. Most heart murmurs can readily be explained on the basis of high velocity flow or abrupt changes in the caliber of the vascular channels.

The Table 3.2 summarizes various types of heart sounds and murmurs with unique perceptual features that distinguish from other murmurs [Leathm, (1975)]. Doctors use these perceptual features (e.g., gushing sound with high pitch, spilt sounds etc.) and play significant role auscultation. The perceptual features carry significant clinical information that can be used for the clinical diagnosis and cardiovascular treatments.

Table 3.3 Classification of murmurs and their behavioral properties.

Murmur Type	**Characteristics**
Systolic ejection	Occurs temporally between S1 and S2. Causes interference to the flow of blood, manifested as turbulence.
Innocent murmurs	Common in young age group and always occur during the systole.
Diastolic murmurs	This murmur occurs at the middle to the end of the diastole and does not allow the laminar passage of blood.

The properties of the heart sounds and murmurs along with their acoustic and behavioral properties are described in the Table 3.2 and Table 3.3 respectively. The Table 3.2 gives a fairs idea about the psychoacoustic properties used by the doctors for the correct interpretation of auscultations.

3.3.2. PRELIMINARIES ON PCG SIGNAL PROCESSING

The PCG signal discloses information about cardiac function through vibrations caused by the working heart. In the early days of PCG signal analysis, manual interpretation of waveform patterns was performed in the time domain. Heart sounds were identified as composite oscillations related to valve closure and heart murmurs seemed to derive from malfunctioning valves or from abnormal holes in the septal wall. When the Fourier transform became practically useful, it provided further information about periodicity and the distribution of signal power. In many biomedical signals, the

Fourier transform showed that sharp frequency peaks were rare, and when they did exist, they often indicated disease [K. Dumont (2004)]. Murmurs possessed characteristics similar to colored noise, and with increasing disease severity, the frequency spectrum became more and more complicated.

In an attempt to disentangle the frequency spectrum, joint time-frequency analysis was employed [Y-T. Zhang, G. Chan, X-Y. Zhan, and L. Yip. (2006)]. In later studies, it could be shown that heart sounds consisted of several components where each component had a main frequency that varied with time. This short introduction basically brings us up to date regarding the tools used for PCG signal analysis. In this thesis, nonlinear techniques will be investigated as means to explore the PCG signal even further. Heart sounds and murmurs are of relatively low intensity and are band-limited to about 10–1000 Hz, (refer to Figure 3.1). Meanwhile the human auditory system, which is adapted to speech, is unable to take in much of this information. An automated signal processing system, equipped with a sound sensor, would be able to exploit this additional information. In a clinical setting, the main tasks for such a system would be to:

- Emphasize the audibility of the PCG signal.
- Extract feature vectors that characterize abnormal events (murmurs) in the PCG signal.
- Extract information suitable for indexing, search and retrieval heart sounds and murmurs.
- Extract information suitable for assessment and classification of heart diseases.

Thus, a PCG processing system interfaced to the PCG databases which contains the recordings of the heart disease and murmurs needs a set of discriminatory feature vectors that assist in the searching, retrieval and classification of cardio vascular disease. The content-based audio algorithms based on similarity measure are more appropriate for many retrieval applications. In the subsequent chapters, the different types content-based algorithms, different types of clustering algorithms based on similarity measures are discussed and also explore the automatic cluster labeling.

CHAPTER – 4

HIERARCHICAL CLUSTERING AND RETRIEVAL OF HEART SOUNDS AND MURMURS

4.0 INTRODUCTION

Pattern Recognition (PR) techniques are extensively used in the Information Retrieval (IR) applications [Salton and McGill (1983)]. These PR techniques are based on statistical decision making or classification using well known statistical procedures such as supervised learning when the training sets are avaliable. In the parametric decision making based on the probability models contain adjustable parameters estimated from the population data. When the distribution parameters are not avaliable or insufficient for the design of the membership function, then it is called the non-parametric decision making techniques [Grossman and Frieder (1998)].

The automatic pattern recognition is an example of supervised learning refers to the process of designing a pattern classifier by using a training set of patterns of known class to determine the choice of a specific decision making technique for classifying the additional similar samples from the unknown. In parametric decision making, samples into divide the naturally occurring groups or clusters based on the measures of similarity without any prior knowledge of class membership.

This chapter explores the hierarchal clustering algorithms using various types of similarity measures and their applicability for the classification and retrieval of audio objects – heart sounds and murmurs. The general principle followed is to exploit the hierarchal structure of the samples and logical hierarchal structure of decision making of the doctors such that the similarity measures those are highest are retrieved and displayed in the result sets of an audio query [Whitman and Rifkin (2002)]. The classical hierarchal clustering algorithms are extended with in-built similarity measures and use them for indexing in effective retrieval of audio objects from the heart sounds and murmurs databases. It is showed that these novel algorithms are useful in audio indexing, searching and retrieval application of heart sounds and murmurs, and evaluate these novel algorithms with the experimental data and discuss the performance evaluation of the proposed algorithms.

4.1 FEATURE SELECTION

Feature selection is the process of selecting a subset of the unique terms occurring in the training set. The subset is used as features in text and audio objects classification and retrieval applications. Feature selection serves two main purposes. First, it makes training and applying a classifier more efficient by decreasing the size of the effective vocabulary. Second, feature selection often increases classification accuracy by eliminating noise features. A *noise feature* is one that, when added to the audio object representation, increases the classification error on new data. Sometimes incorrect generalization of the training set leads to over-fitting which impairs the quality of the search results [Byrd (2010)]. It can be viewed as a feature selection mechanism for replacing a complex classifier (using all features) with simple ones.

```
SELECT-FEATURES (D, c, k)
V ← Extarct_Features (D)                                /* extract feature vectors */
L ← []
for each t ε V
  do A (t, c) ← Computer_Feature_Utility (D, t, c)    /*get relevant features */
     Append (L, <A (t, c), t>)
return Features_List_Largest_Value (L, k)
```

Figure 4.1 Feature selection algorithm for audio retrieval.

The basic feature selection algorithm is described in Figure 4.1. For a given class c, we compute a utility measure $A(t, c)$ for each term of the vocabulary or feature vectors and select the k terms that have the highest values of $A(t, c)$. All other terms are discarded and not used in classification. Three different utility measures: mutual information, $A(t, c) = I(U_t; C_c)$; the X^2 test, $A(t, c) = X^2(t, c)$; and frequency, $A(t, c) = N(t, c)$. Similar reasoning holds true for the multi-class classification with multiple feature vectors.

4.2 HIERARCHICAL CLUSTERING - HC

Flat clustering is efficient and conceptually simple; it has a number of drawbacks. The flat unstructured set of clusters, require a prespecified number of clusters as input and are nondeterministic. Hence their non-deterministic behaviors are unreliable and not effective in retrieval applications [Chen and Lu, (1996)].

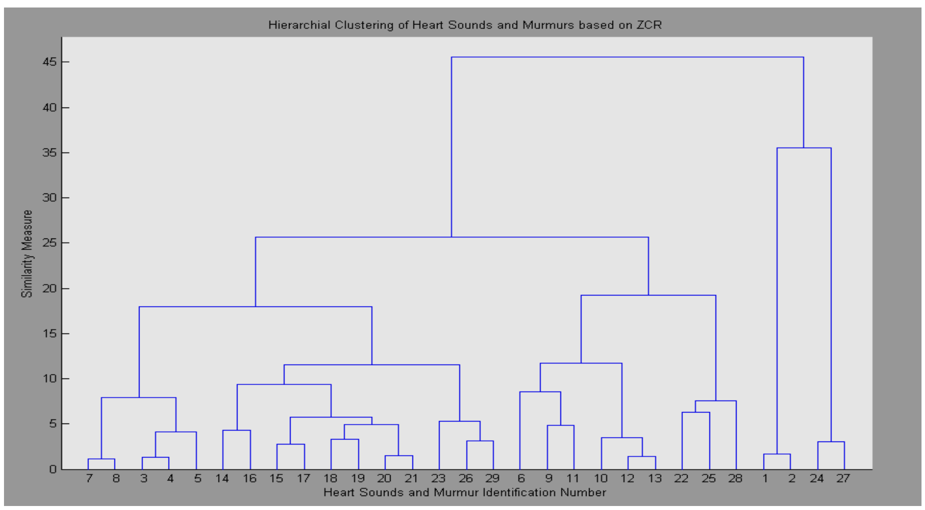
Hierarchical Clustering (HC) techniques outputs a hierarchy, a structure that is more informative than the unstructured set of clusters returned by flat clustering. Hierarchical clustering does not require us to prespecify the number of clusters and are deterministic in nature. The deterministic behavior of hierarchical clustering comes at the cost of lower efficiency. The hierarchical clustering algorithms have a complexity that is at least quadratic in the number of audio objects compared to the linear complexity of *K*-means and Euclidian Distance Measures - EDM. This chapter introduces *agglomerative* hierarchical clustering and presents four different agglomerative algorithms, which differ in the similarity measures they employ*: single-link, complete-link, group-average, and centroid similarity* and discuss their applicability for the retrieval of heart sounds and murmurs as audio objects from the heart sound and murmur database. It also discusses the introduction of the novel features (psychoacoustic features) for the automatic classification and retrieval of audio objects based on various similarity measures [B. Logan and Salomon (2001)].

It covers the optimality conditions of hierarchical clustering for classification and retrieval applications based on audio feature sets based on time and frequency domains as well as psychoacoustic features. The automatic labeling of clusters needs to be addressed whenever humans interact with the output of clustering. Automatic labeling is useful in the design and user interfaces and query evaluation. The implementation issues and specific extensions for audio classification and retrieval applications are discussed. In particular, hierarchical clustering is appropriate for any of the pattern recognition and information retrieval applications when the index is dynamically generated by analyzing the audio query. The audio query used by the doctors is of the type query by example (QBE), in which a sample PCG audio recording is supplied to the classfier; the features are derived and modeled in the vector space. The similarity measures, i.e., the sample distance, cluster distance, inter-cluster distance, cluster centroid, average cluster centroid are used for the membership function as well decisions boundary function. In the user interface, the result sets are ranked and played for the doctor for making effective clinical diagnosis. The PCG recordings are also plotted on the screen as a function of time and support visual inspection.

4.2.1 HIERARCHICAL AGGLOMERATIVE CLUSTERING - HAC

Hierarchical clustering algorithms are either top-down or bottom-up approach for the traversal of index (for example, B tree) tree. Bottom-up algorithms treat each audio object as a singleton cluster at the outset and then successively merge (or *agglomerate*) pairs of clusters until all clusters have been merged into a single cluster that contains all audio objects. Hence, bottom-up hierarchical clustering is therefore called *hierarchical agglomerative cluster - HAC and* more popular in IR application than top-down algorithm [Pampalk (2004)].

Similarity
Measure



PCG identification number

Figure 4.2 Dendrogram of heart diseases and murmurs with identification numbers based ZCR as feature vector.

Before looking at specific similarity measures used in HAC, first introduce a method for depicting hierarchical clusterings graphically, discuss a few key properties of HACs and present a simple algorithm for computing an HAC and show how it can be used for the audio retrieval applications. It is recommended to use various discriminatory features vectors for the indexing, searching and classifications of heart sounds and murmurs. A typical feature vector – Zero Crossing Rate (ZCR), the rate of change in sign i.e. positive to negative or negative to positive is closely related to the frequency of the signal. The ZCR is extensively studied and used in speech research. For example, it is used for the separation of speech from the pause in a speech recording. In PCG research, it is mainly used for the study of heart murmurs as they have noise-like features and can be used as discriminatory feature for murmurs retrieval and heart disease classifications. For experimental purpose, a sample of thirty heart sounds and murmurs stored and assigned unique PCG identification (PCG ID) and calculated the ZCR for each heart diseases and murmurs as shown in table 4.1 (a), (b) and (c) and show a high variations among the representative heart diseases and murmurs.

Table 4.1 The PCG recordings (a), (b) and (c) of heart diseases and murmurs with diseases identification number for the feature vector - ZCR.

PCG ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
ZCR	**37.74**	**38.39**	**3.854**	**4.135**	**6.606**	**9.23**	**29.0**	**1.532**	**1.358**	**25.89**	**18.16**	**23.07**

(a)

PCG ID	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
ZCR	**17.37**	**16.96**	**10.28**	**2.125**	**7.968**	**1.358**	**1.358**	**4.311**	**6.59**	**2.792**	**20.52**	**6.602**

(b)

PCG ID	**25**	**26**	**27**	**28**	**29**	**30**
ZCR	**48.56**	**17.20**	**5.826**	**48.56**	**21.91**	**5.744**

(c)

Table 4.2 The normalized values of similarity measures in the dendrogram.

Absolute Value	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**
Normalized Value	**0.0**	**0.11**	**0.22**	**0.33**	**0.44**	**0.55**	**0.66**	**0.77**	**0.88**	**1.0**

An HAC clustering is typically visualized as a *dendrogram* as shown in Figure 4.2. The ZCR data from the table 4.1 (a) (b) and (c) is used to draw the dendrogram for various heart and murmurs diseases. After finding the similarity measure among the clusters and merged by a horizontal line as they represent the similar measure of similarity. Each merge is represented by a horizontal line. For example, we have a horizontal line and it intersects the y-coordinate at an absolute value of 35, measure of similarity and when normalized, its value is about (0.77) using the Table 4.2. With similar reasoning, we can observe that the heart and murmurs disease with identification number 1and 2 are closely related and similarly 24 and 27 PCG identification numbers. The y-coordinate of the horizontal line (refer fig. 4.2) of the two clusters that were merged, where audio objects are viewed as singleton clusters and represents the similarity measure. This similarity is called the *combination similarity* of the merged cluster and defined as combination similarity of a singleton cluster as its document's self-similarity (which is 1.0 for cosine similarity).

By moving up from the bottom layer to the top node, a dendrogram allows us to reconstruct the history of merges that resulted in the depicted clustering with a fundamental assumption in HAC is that the merge operation is *monotonic*. Monotonic means that if s_1, $s_2 \ldots s_{K-1}$ are the combination similarities of the successive merges of an HAC, then $s_1 \geq s_2 \geq \ldots \geq s_{K-1}$ holds. A non-inversion monotonic hierarchical clustering contains at least one *inversion* $s_i < s_{i+1}$ and contradicts the fundamental assumption that the best merge is available at each step. Hierarchical clustering does not require a prespecified number of clusters. However, in some applications, it is better to partition of disjoint clusters just as in flat clustering. In those cases, the hierarchy needs to be cut at some point and can be used in the user interface to the audio retrieval applications [Charikar et el., (1997)]].

In general, the following criteria can be used to determine the cutting point:

• Cut at a prespecified level of similarity. In Figure 4.2, for example, we cut the dendrogram at 0.4 if we want clusters with a minimum combination similarity of 0.4. In Figure 4.2, cutting the diagram at $y = 0.4$ yields 18 clusters (grouping only audio objects with high similarity together). The cut operation can be implemented in the user interface and iteratively the doctors can interact with the dendrogram and derive the best match to a given audio query.

• Cut the dendrogram where the gap between two successive combination similarities is largest. Such large gaps arguably indicate "natural" clusterings. Adding one more cluster decreases the quality of the clustering significantly, so cutting before this steep decrease occurs is desirable.

$$K = \underset{K'}{\arg\min} \, [\, \mathrm{RSS}\,(K') + \lambda K' \,] \tag{4.1}$$

where K' refers to the cut of the hierarchy that results in K' clusters, RSS is the residual sum of squares and λ is a penalty for each additional cluster.

• As in flat clustering, it is possible to prespecify the number of clusters K and select the cutting point that produces K clusters.

4.2.2 DESIGN OF HAC ALGORITHM FOR CLASSIFICATION AND RETRIEVAL

A simple HAC algorithm is described in Figure 4.3. We first compute the $N \times N$ similarity matrix C. The algorithm then executes $N - 1$ steps of merging the currently most similar clusters. In each iteration, the two most similar clusters are merged and the rows and columns of the merged cluster i in C are updated. The clustering is stored as a list of merges in A. I indicate which clusters are still available to be merged [Wu and Manber (1996)].

The function SIM (i, m, j) computes the similarity of cluster j with the merge of clusters I and m. For some HAC algorithms, SIM (i, m, j) is simply a function of $C[j]$ $[i]$ and $C[j]$ $[m]$, for example, the maximum of these two values for single-link. The algorithm is refined for the different similarity measures of single-link and complete-link clustering and group-average and centroid clustering based on similarity distance measures for the classification and retrieval of hearts sounds and murmurs.

```
SIMPLE-HAC (d1. . . dN)

for n ← 1 to N
  do for i ← 1 to N
  do C[n] [i] ← SIM (dn, di)   /* find the similarity measure */
  I[n] ← 1                     /* keeps track of active clusters */
  A ← []       /* assembles clustering as a sequence of merges*/
    for k ← 1 to N − 1
      do <i, m> ← argmax {(i, m: i≠m^I[i] =1^I[m] =1} C[i][m]
           A.APPEND (<i, m>)          /* store and merge*/
         for j ← 1 to N
        do C[i][j] ← SIM(i,m, j)
              C[j][i] ← SIM(i,m, j)
              I[m] ← 0         /* deactivate cluster */
 return A
```

Figure 4.3 A simple HAC algorithm.

4.2.3 TIME COMPLEXITY OF HAC

In this section, the time complexity of the simple HAC algorithm is Θ (N^3) (refer to the Figure 4.3) because it exhaustively scan the $N \times N$ matrix C for the largest similarity in each of $N - 1$ iterations. It is found that, a more efficient algorithm is the priority-queue algorithm (shown in Figure 4.7). Its time complexity is Θ ($N^2 \log N$). The rows $C[k]$ of the $N \times N$ similarity matrix C are sorted in decreasing order of similarity in the priority queues P. $P[k]$.MAX() then returns the cluster in $P[k]$ that currently has the highest similarity with ω_k, where we use ω_k to denote the k^{th} cluster. After creating the merged cluster of ω_{k1} and ω_{k2}, ω_{k1} is used as its representative. The function SIM computes the similarity function for potential merge pairs: largest similarity for single-link clustering, smallest similarity for complete-link clustering, and average similarity for group-average agglomerative cluster - GAAC and centroid similarity for centroid clustering. These time and space complexity shed light on the query performance and has significant impact on the indexing, index computation and ranking algorithms.

4.3 SINGLE-LINK AND COMPLETE-LINK CLUSTERING

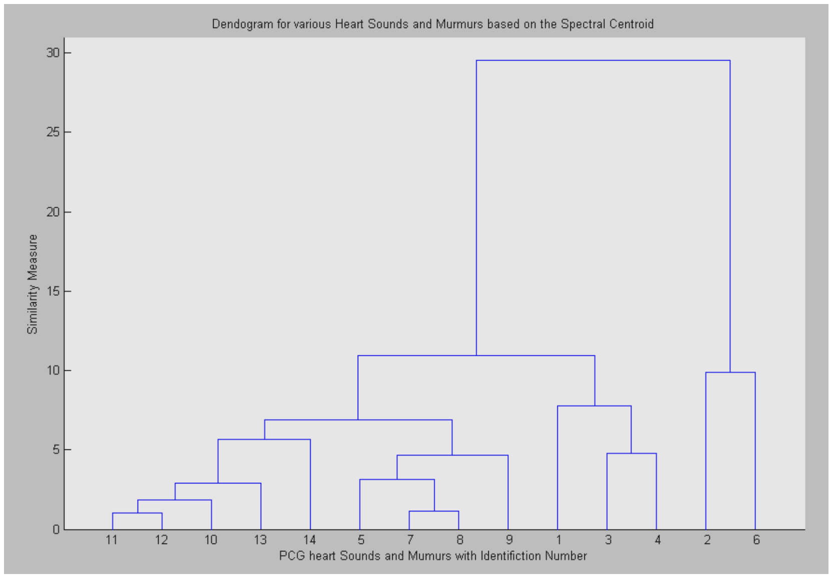
In *single-link clustering or single-linkage clustering*, the similarity of two clusters is the similarity of their *most similar* members. This single-link merge criterion is *local* and focuses on to the area where the two clusters come closest to each other and affect the quality of the query.

In *complete-link clustering* or *complete-linkage clustering*, the similarity of two clusters is the similarity of their *most dissimilar m*embers and this is equivalent to choosing the cluster pair whose merge has the smallest diameter. This complete-link merge criterion is non-local; the entire structure of the clustering can influence merge decisions. This results in a preference for compact clusters with small diameters but also causes sensitivity to outliers. A single document or audio object far from the center can increase diameters of candidate merge clusters dramatically and completely change the final clustering.

For demonstration purpose, it is possible to visualize a single-link and a complete-link clustering of fourteen audio objects of the heart sounds and murmurs [Volker and Gunter (1998)]. The first four steps, each producing a cluster consisting of a pair of two documents, are identical. Then single-link clustering joins the upper two pairs (and after that the lower two pairs) because on the maximum similarity definition of cluster similarity, those two clusters are closest. The local criterion in single-link clustering can cause undesirable elongated clusters. Complete link clustering joins the left two pairs (and then the right two pairs) because those are the closest pairs according to the minimum-similarity definition of cluster similarity.

Both single-link and complete-link clustering have graph-theoretic interpretations. Define s_k to be the combination similarity of the two clusters merged in step k, and $G\ (s_k)$ the graph that links all data points with a similarity of at least s_k. Then the clusters after step k in single-link clustering are the connected components of $G\ (s_k)$ and the clusters after step k in complete-link clustering are maximal cliques of $G\ (s_k)$.

Similarity
measures



PCG identification number

Figure 4.4 Dendrogram of fourteen heart diseases with identification numbers based on feature vector - spectral centroid.

A *connected component* is a maximal component set of connected points such that there is a path connecting each pair. A *clique* is a set of points that are completely linked with each other. These graph-theoretic interpretations motivate the terms single-link and complete-link clustering. Single-link clusters at step k are maximal sets of points that are linked via at least one link (a single link) of similarity $s \geq s_k$; complete-link clusters at step k are maximal sets of points that are completely linked with each other via links of similarity $s \geq s_k$.

Single-link and complete-link clustering reduce the assessment of cluster quality to a single similarity between a pair of documents or audio objects: the two most similar documents in single-link clustering and the two most dissimilar documents in complete-link clustering. A measurement based on one pair cannot fully reflect the distribution of documents in a cluster. It is therefore not surprising that both algorithms often produce undesirable clusters. Single-link clustering can produce straggling clusters, since the merge criterion is strictly local, a chain of points can be extended for long distances without regard to the overall shape of the emerging cluster and is called *chaining*. The chaining effect is also apparent in Figure 4.4.

Table 4.3 Sample heart diseases with identification numbers and murmurs based on feature - spectral centroid of the PCG recordings.

PCG ID	Heart Disease and Mummers	Spectral Centroid
1	Systolic-Mitral-Prolapse	2.22E+00
2	Systolic Split S1	3.08E+01
3	Systolic Mitral Regugration	5.63E+00
4	Systolic Pulmonary Stenosis	9.41E+00
5	Systolic Ventricular Septal Defect	3.42E+00
6	Systolic Mitral Value Replacement	2.49E+01
7	Normal Heart	2.85E+00
8	Aortic Stenosis	2.70E+00
9	Mitral stenosis	5.34E+00
10	Distolic -Ventricular Septal Defect	1.22E+00
11	4th Heart Sound	8.37E-01
12	3rd Heart Sound	8.64E-01
13	Diastolic-Rumble	1.85E+00
14	Ejection-Click	4.34E+00

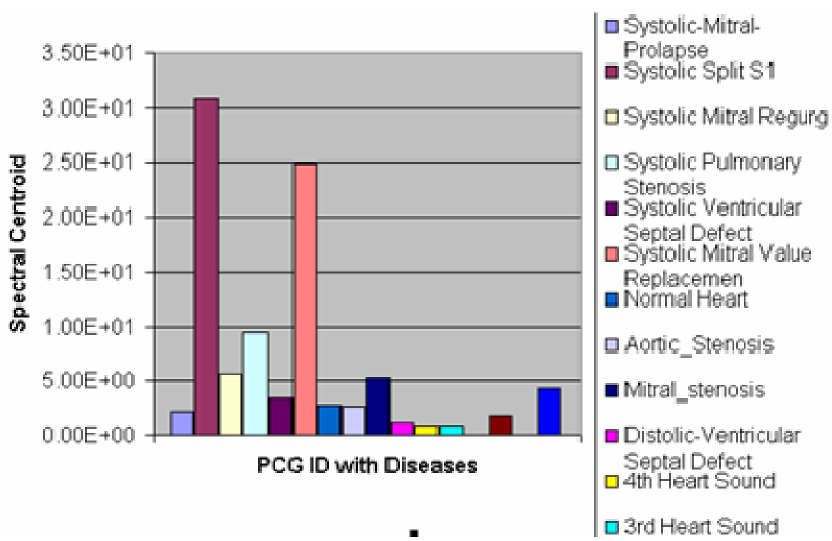


Figure 4.5 Visualization of sample heart diseases with identification numbers based on feature - spectral centroid.

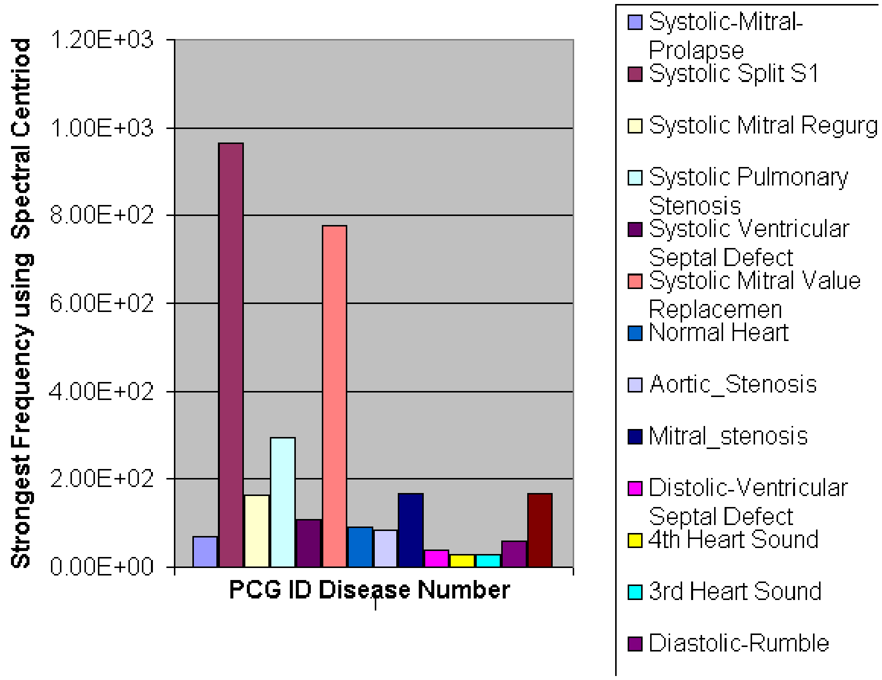


Figure 4.6 Visualization of sample heart diseases with identification numbers based on strongest frequency via - spectral centroid.

The last eleven merges of the single-link clustering (those above the 0.1 line) add on single documents or pairs of documents, corresponding to a chain. The complete-link clustering in shown in Figure 4.4 and avoids this problem. Documents are split into two groups of roughly equal size when we cut the dendrogram at the last merge. In general, this is a more useful organization of the data than a clustering with chains. However, complete-link clustering suffers from a different problem. It pays too much attention to outliers, points that do not fit well into the global structure of the cluster.

Table 4.4 Sample heart diseases and murmurs with identification numbers based on strongest frequency via - spectral centroid.

PCG ID	Heart Disease and Mummers	Highest Frequency via Spectral Centroid
1	Systolic-Mitral-Prolapse	6.93E+01
2	Systolic Split S1	9.63E+02
3	Systolic Mitral Regugration	1.62E+02
4	Systolic Pulmonary Stenosis	2.94E+02
5	Systolic Ventricular Septal Defect	1.07E+02
6	Systolic Mitral Value Replacement	7.79E+02
7	Normal Heart	8.89E+01
8	Aortic Stenosis	8.42E+01
9	Mitral stenosis	1.67E+02
10	Distolic -Ventricular Septal Defect	3.81E+01
11	4th Heart Sound	2.62E+01
12	3rd Heart Sound	2.70E+01
13	Diastolic-Rumble	5.79E+01
14	Ejection-Click	1.67E+02

As an example of how a row of C is processed the loop in lines 1–7 has time complexity Θ (N^2) and the loop in lines 9–21 has time complexity $\Theta(N^2 \log N)$ for an implementation of priority queues that supports deletion and insertion in Θ (log N). The overall complexity of the algorithm is therefore Θ (N^2 log N). In the definition of the function SIM, $\sim v_m$ and $\sim v_i$ are the vector sums of $\omega_{k1} \mathrm{U} \omega_{k2}$ and ω_i, respectively, and N_m and N_i are the number of audio objects in $\omega_{k1} \mathrm{U} \omega_{k2}$ and ω_i, respectively.

```
EFFICIENT-HAC-INDEXING (~d1. . . ~dN)
for n ← 1 to N
  do for i ← 1 to N
       do C[n][i].sim ← ~dn · ~di
          C[n][i].index ← i
          I[n] ← 1
          P[n] ← priority queue for C[n]      /* sorted on sim */
          P[n].DELETE(C[n][n])
          A ← []
               for k ← 1 to N −1
                    do k1 ← argmax {k: I[k] =1} P[k].MAX ().sim
                         k2 ← P [k1].MAX ().index
                         A.APPEND (<k1, k2>)
                         I[k2] ← 0
                         P[k1] ← []
               for each i with I[i] = 1 ^ i ≠ k1
               do P[i].DELETE(C[i][k1])
                    P[i].DELETE(C[i][k2])
                    C[i][k1].sim ← SIM(i, k1, k2)
                    P[i].INSERT(C[i][k1])
                    C[k1][i].sim ← SIM(i, k1, k2)
                    P[k1].INSERT(C[k1][i])
return A
```

Figure 4.7 An efficient HAC algorithm with priority queue and indexing.

In single-link clustering algorithm uses a nearest best merge - NBM array for storing intermediate results. After merging two clusters i_1 and i_2, the first one (i_1) represents the merged cluster. If $I[i] = i_1$, then I is the representative of its current cluster. If $I\ [i] = i_2$, then i has been merged into the cluster represented by $I\ [i]$ and will therefore be ignored when updating a nearest best merge array $[i_1]$.

```
SINGLE-LINK-CLUSTERING (d_1, . . . , d_N)
1     for n ← 1 to N
2           do for i ← 1 to N
3           do C[n][i].sim ← SIM(d_n, d_i)
4                 C[n][i].index ← i
5                 I[n] ← n
6                 NBM[n] ← argmaxX ε{C[n] [i]:n≠i} X.sim
7                 A ← []
8                 for n ← 1 to N − 1
9                 do i_1 ← argmax {i:I[i]=i} NBM[i].sim
10                 i_2 ← I [NBM[i1].index]
11                A.APPEND (<i_1, i_2>)
12                for i ← 1 to N
13                do if I[i] = i ^ I ≠ i1 ^ i≠ i_2
14                  then C[i_1][i].sim ← C[i][i_1].sim ← max(C[i_1][i].sim, C[i2][i].sim)
15                if I[i] = i_2
16                 then I[i] ← i_1
17                NBM[i_1] ← argmaxX ε {C[i_1][i]:I[i]=I ≠i_1} X.sim
18    return A
```

Figure 4.8 Single-link clustering algorithm with indexing.

The argument of EFFICIENT-HAC-INDEX algorithm in Figure 4.7 is a set of vectors because GAAC and centroid clustering require vectors as input. The complete-link version of EFFICIENT-HAC-INDEX can also be applied to documents that are not represented as vectors. For single-link, we can introduce a next-best-merge array (NBM) as a further NBM keeps track of what the best merge is for each cluster. In other words, the best-merge candidate for the merged cluster is one of the two best-merge candidates of its components in single-link clustering. This means that C can be updated in $\Theta(N)$ in each iteration – by taking a simple maximum of two values as shown on line 14 in Figure 4.7 for each of the remaining $\leq N$ clusters. This is because the complete-link merge criterion is non-local and can be affected by points at a great distance from the area where two merge candidates meet. In practice, the efficiency penalty of the $\Theta(N^2 \log N)$ algorithm is small compared with the $\Theta(N^2)$ single-link algorithm since computing the similarity between two audio objects (e.g., as a dot product) is an order of magnitude slower than comparing two scalars in sorting. It may be observed that, the HAC algorithms have time complexity Θ (N^2) with respect to similarity computations [Monolopoulos et el., (1994)].

4.4 GROUP-AVERAGE AGGLOMERATIVE CLUSTERING - GACC

Group-average agglomerative clustering evaluates cluster quality based on *all* similarities between documents, thus avoiding the pitfalls of the single-link and complete-link criteria, which equate cluster similarity with the similarity of a single pair of documents. GAAC is also called *group-average clustering* and *average-link clustering*. GAAC computes the average similarity SIM-GA of all pairs of documents or audio objects, including pairs from the same cluster [Roussopoulos et el., (1995)]. But self-similarities are not included in the average:

$$\text{SIM-GA}(\omega_i, \omega_j) = \frac{1}{(N_i + N_j)(N_i + N_j - 1)} \sum_{d_m \in \omega_i \cup \omega_j} \sum_{d_n \in \omega_i \cup \omega_j, d_n \neq d_m} \vec{d}_m \cdot \vec{d}_n \tag{4.1}$$

- denotes the dot product, and N_i and N_j are the number of documents in ω_i and ω_j,

respectively.

The motivation for GAAC is that in selecting two clusters ωi and ωj as the next merge in HAC is that the resulting merge cluster $\omega k = \omega i$ U ωj should be coherent. To judge the coherence of ω_k, it is appropriate to look at all document-document similarities within ω_k, including those that occur within ω_i and those that occur within ω_j. It is possible to compute the measure SIM-GA efficiently because the sum of individual vector similarities is equal to the similarities of their sums:

$$\sum_{d_m \in \omega_i} \sum_{d_n \in \omega_j} (\vec{d}_m \cdot \vec{d}_n) = (\sum_{d_m \in \omega_i} \vec{d}_m) \cdot (\sum_{d_n \in \omega_j} \vec{d}_n) \tag{4.2}$$

$$\text{SIM-GA}(\omega_i, \omega_j) = \frac{1}{(N_i + N_j)(N_i + N_j - 1)} [(\sum_{d_m \in \omega_i \cup \omega_j} \vec{d}_m)^2 - (N_i + N_j)]$$

The term $(N_i + N_j)$ on the right is the sum of $N_i + N_j$ self-similarities of value 1.0. With this approach, it is possible to compute cluster similarity in constant time assuming that we have available the two vector sums instead of in $\Theta(N_i N_j)$.

$$\sum_{d_m \in \omega_i} \vec{d}_m \text{ and } \sum_{d_m \in \omega_i} \vec{d}_m) \qquad (4.3)$$

This is important because we need to be able to compute the function SIM on lines 18 and 20 in EFFICIENT-HAC-INDEX (refer figure 4.7) in constant time for efficient implementations of GAAC. Note that for two singleton clusters, equation (4.3) is equivalent to the dot product. Equation (4.2) relies on the distributives of the dot product with respect to vector addition. Since this is crucial for the efficient computation of a GAAC clustering, the method cannot be easily applied to representations of documents that are not real-valued vectors. Also, equation (4.2) only holds for the dot product. It is observed that these algorithms have near-equivalent descriptions in terms of dot product, cosine similarity and Euclidean distance equation (4.2) can only be expressed using the dot product. This is a fundamental difference between single-link/complete-link clustering and GAAC. The first two only require a square matrix of similarities as input and ignore how these similarities were computed.

To summarize, GAAC requires

(i) Documents or audio objects represented as vectors

(ii) Length normalization of vectors, so that self-similarities are 1.0

(iii) The dot product as the measure of similarity between vectors and sums of vectors.

The merge algorithms for GAAC and complete-link clustering are the same except that we use equation (4.3) as similarity function in Figure 4.7. Therefore, the overall time complexity of GAAC is the same as for complete-link clustering: Θ (N^2 log N). Like complete-link clustering, GAAC is not best merge persistent, this means that there is no Θ (N^2) algorithm for GAAC that would be analogous to the Θ (N^2) algorithm for single-link

We can also define group-average similarity as including self-similarities:

$$\text{SIM-GA}'(\omega_i, \omega_j) = \frac{1}{(N_i + N_j)^2} \Big(\sum_{d_m \in \omega_i \cup \omega_j} \vec{d}_m \Big)^2 = \frac{1}{N_i + N_j} \sum_{d_m \in \omega_i \cup \omega_j} [\vec{d}_m \cdot \vec{\mu}(\omega_i \cup \omega_j)] \qquad (4.4)$$

where the centroid ~$\mu(\omega)$ is defined and is equivalent to the intuitive definition of cluster quality as average similarity of documents ~d_m to the cluster's centroid ~μ. Self-similarities are always equal to 1.0, the maximum possible value for length normalized vectors. The proportion of self-similarities in equation (4.4) is $i/i^2 = 1/i$ for a cluster of size i. This gives an unfair advantage to small clusters since they will have proportionally more self-similarities. For two audio objects d_1, d_2 with a similarity s, we have SIM-GA′ $(d_1, d_2) = (1 + s)/2$. In contrast, SIM-GA $(d_1, d_2) = s \leq (1 + s)/2$. This similarity SIM-GA (d_1, d_2) of two audio objects is the same as in single-link, complete-link and centroid clustering. It is preferred to use equation (4.3), which excludes self-similarities from the average; because it is better not to penalize large clusters or their smaller proportion of self-similarities and it is necessary to have a consistent similarity value s for document pairs in all HAC algorithms.

4.5 CENTROID CLUSTERING

In centroid clustering, the similarity of two clusters is defined as the similarity of their centroids:

$$\begin{aligned}
\text{SIM-CENT}(\omega_i, \omega_j) &= \vec{\mu}(\omega_i) \cdot \vec{\mu}(\omega_j) \\
&= \Big(\frac{1}{N_i} \sum_{d_m \in \omega_i} \vec{d}_m\Big) \cdot \Big(\frac{1}{N_j} \sum_{d_n \in \omega_j} \vec{d}_n\Big) \qquad (4.5) \\
&= \frac{1}{N_i N_j} \sum_{d_m \in \omega_i} \sum_{d_n \in \omega_j} \vec{d}_m \cdot \vec{d}_n
\end{aligned}$$

The above equation gives centroid similarity and shows that centroid similarity is equivalent to average similarity of all pairs of documents from *different* clusters. Thus, the difference between GAAC and centroid clustering is that GAAC considers all pairs of documents in computing average pair wise similarity whereas centroid clustering excludes pairs from the same cluster [Grimaldi et el., (2003)]. Like GAAC, centroid clustering is not best-merge persistent and therefore Θ (N^2 log N). In contrast to the other three HAC algorithms, centroid clustering is not monotonic and *inversions* can occur: Similarity can increase during hierarchical clustering. The non-monotonic inversion in the hierarchical clustering of the three points appears as an intersecting merge line in the dendrogram. This is an example of an inversion: similarity *increases* in this sequence of two clustering steps. In a monotonic HAC algorithm, similarity is monotonically *decreasing* from iteration to iteration. Increasing similarity in a series of HAC clustering steps contradicts the fundamental assumption that small clusters are more coherent than large clusters. An inversion in a dendrogram shows up as a horizontal merge line that is *lower* than the previous merge line state the optimality conditions of hierarchical clustering precisely, we first define the combination similarity COMB-SIM of a clustering W = $\{\omega_1, \ldots, \omega_K\}$ as the smallest combination similarity of any of its K clusters:

$$\text{COMB-SIM}(\{\omega_1, \ldots, \omega_K\}) = \min k\ \text{COMB-SIM}(\omega_k)$$

Recall that the combination similarity of a cluster ω that was created as the merge of ω_1 and ω_2 is the similarity of ω_1 and ω_2. We then define W = $\{\omega_1, \ldots, \omega_K\}$ to be *optimal* if all clusterings W′ with k clusters, $k \leq K$, have lower combination similarities: $|W'| \leq |W| \Rightarrow$ COMB-SIM(W′) $\leq$ COMB-SIM(W).Centroid clustering is not optimal because inversions can occur. The above definition of optimality would be of limited use if it was only applicable to a clustering together with its merge history. If we use these definitions of combination similarity, then optimality is a property of a set of clusters and not of a process that produces a set of clusters.

The inductive basis of the proof is that a clustering with $K = N$ clusters has combination similarity 1.0, which is the largest value possible. The induction hypothesis is that a single-link clustering W_K with K clusters is optimal: COMB-SIM(W_K) $\geq$ COMB-SIM(W'_K) for all W_K. Assume for contradiction that the clustering W_{K-1} we obtain by merging the two most similar clusters in W_K is not optimal and that instead a different sequence of merges W'_K, W'_{K-1} leads to the optimal clustering with $K - 1$ clusters. We can write the assumption that W'_{K-1} is optimal and W_{K-1} is not as COMB-SIM(W'_{K-1}) > COMB-SIM(W_{K-1}). It is recommended using GAAC for document clustering because it is generally the method that produces the clustering with the best method combination similarity time complexity. Comment single-link max inter-similarity of any two documents or audio objects $\Theta(N^2)$ yes chaining effect complete-link min inter-similarity of any two documents or audio objects $\Theta(N^2 \log N)$ not sensitive to outliers group-average average of all sims $\Theta(N^2 \log N)$ not best choice for most applications centroid average inter-similarity $\Theta(N^2 \log N)$. It does not suffer from chaining, from sensitivity to outliers and from inversions. Again, the decision whether a group of documents are duplicates of each other is not influenced by audio objects that are located far away and single-link clustering is a good choice for duplicate detection.

4.6 CLUSTER LABELING

In many applications [Hu and Dannenberg (2002), Kosugi et el., (1999)] of flat clustering and hierarchical clustering, particularly in analysis tasks and in user interfaces human users interact with clusters. In such settings, we must label clusters, so that users can see what a cluster is about. *Differential cluster labeling* selects cluster labels by comparing the distribution of terms in one cluster with that of other clusters. In particular; mutual information (MI) will identify cluster labels that characterize one cluster in contrast to other clusters. A combination of a differential test with a penalty for rare terms often gives the best labeling results because rare terms are not necessarily representative of the cluster as a whole. By apply three labeling methods to a *K*-means clustering and is observed that there is almost no difference between MI and $\chi 2$. Hence, *Cluster-internal labeling* computes a label that solely depends on the cluster itself, not on other clusters.

In [Jain and Dubes (1998)] labeling a cluster with the metadata (e.g., patient case history, PCG disease identification number) closest to the centroid is one cluster-internal method. The cluster summaries computed by three labeling methods: most highly weighted terms in centroid (centroid), mutual information, and the patient case history, closest to the centroid of the cluster (title). However, a single document or audio object is unlikely to be representative of all documents in a cluster-internal methods are efficient, but they fail to distinguish terms that are frequent in the collection. It is possible to derive a good sense of the documents or audio objects in a cluster from scanning the selected terms. For hierarchical clustering, additional complications arise in cluster labeling. It is necessary to distinguish an internal node in the tree from its siblings, but also from its parent and its children.

Documents or audio objects in child nodes are by definition also members of their parent node, so we cannot use a naive differential method to find labels that distinguish the parent from its children. However, more complex criteria, based on a combination of overall collection frequency and prevalence in a given cluster, can determine whether a term is a more informative label for a child node or a parent node.

CHAPTER – 5

PSYCHOACOUSTICS MODELING AND RETRIEVAL OF HEART SOUNDS

5.0 INTRODUCTION

Psychoacoustics is the science in which we quantify the human perception of sounds [E. Zwicker and H. Fastl, (1999)]. It is possible to establish some objective variables that will be useful for the human perception of sounds. For frequency and intensity, standardized instruments can produce the outputs that are linearly proportional to the stimulus. For example, we can count the zero crossings of a sinusoid over a prescribed interval can be calibrated to read and can used to define the frequency of the signal. Another motivating example, a measure of the spectrum of a sound can also be defined, for instance by particular form of spectrogram. Duration is another objective property of a sound. Each of these sound characteristics has a corresponding perceptual variable. The perception of frequency is called pitch, the perception of intensity is called loudness and perception of spectrum is called timbre. These human response variables are not linearly proportional to the value of the corresponding stimulus variables. Thus, if a person hears a pure tone at some given frequency f, followed by another tone at 2f, the perception will not be at that frequency of the second tone is double that of the first one. Furthermore, the response variables are often dependent on more than one stimulus variables. For instance, the subjective impression of pitch, although primarily dependent on frequency can vary other parameters, such as intensity, spectrum, loudness and timbre. In psychoacoustic research it found that auditory sensations increase logarithmically as the intensity of the stimulus increases. To define a subjective measure of loudness, we have to introduce the sone as per the work based on the [E. Zwicker and H. Fastl, (1999)] and empirical relations between the sound pressure p and the loudness S in sones gives the result

$$S \;\alpha\; p^{0.6} \tag{5.1}$$

where a sone value of one is set to be the loudness of a 1000 Hz tone at an intensity of a 40-dB sound pressure level (SPL). We know that the sound intensity is proportional to the square of the pressure, hence we obtain

$$S \ \alpha \ I^{0.3} \tag{5.2}$$

In general, w can say that loudness is proportional to the cube root of the intensity. We also that, the difference between sound energy levels is measured in decibels

$$L = 10 \log_{10} I_1/I_2 \tag{5.3}$$

Since the intensity is proportional to the square of the sound pressure,

$$L = 20 \log_{10} p_1/p_2 \tag{5.4}$$

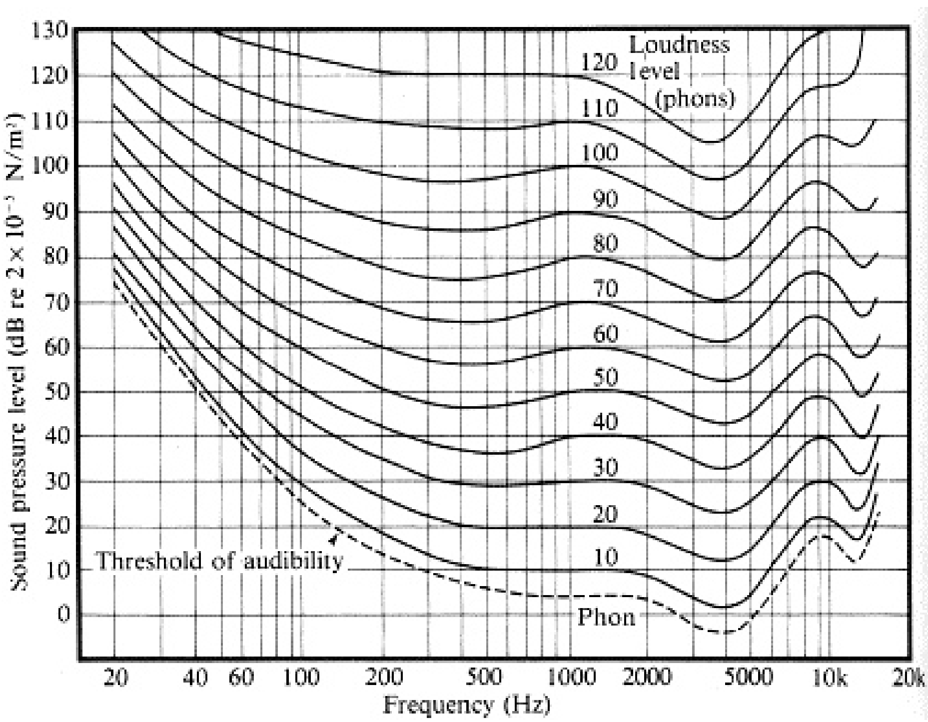


Figure 5.1 Fletcher-Munson curves or equal-loudness contour standards.

An equal-loudness contour is a measure of sound pressure in dB or SPL over the frequency spectrum, for which a listener perceives a constant loudness when presented with pure steady tones. The unit of measurement for loudness levels is the phone and is arrived at by reference to equal-loudness contours. Equal-loudness contours are often referred to as "Fletcher-Munson"' curves, after the earliest researchers. [E. Zwicker and H. Fastl, (1999)], but those studies have been superseded and incorporated into newer standards such as ISO 226:2003.

The ear is not equally sensitive to all frequencies, particularly in the low and high frequency ranges. The response to frequencies over the entire audio range has been charted, originally by Fletcher and Munson in 1933, with later revisions by other authors, as a set of curves showing the sound pressure level of pure tones that are perceived as being equally loud. The curves are plotted for each 10 dB rise in level with the reference tone being at 1 KHz. The curves are lowest in the range from 1 to 5 KHz, with a dip at 4 KHz, indicating that the ear is most sensitive to frequencies in this range. The intensity level of higher or lower tones must be raised substantially in order to create the same impression of loudness. The phon scale was devised to express this subjective impression of loudness, since the decibel scale alone refers to actual sound pressure or sound intensity levels. In the audio retrieval research, the Mel Frequency Ceptral Coefficient (MFCC), 13-dimension vector characterizes psychoacoustic properties of the audio perception [Logan (20000]. In this thesis work, the MFCC is used as a discriminatory feature vector and used it for heart sound retrieval algorithms and applications. A content based retrieval of heart sounds based on (MFCC) is explored in the subsequent discussions. Histogram modeling using MFCC feature vectors and various techniques of similarity measures are discussed for pattern matching of PCG signals. Search accuracy of histogram matching is tested with respect to correlation algorithms and it is observed that histogram algorithm for content based heart sound retrieval is more efficient and accurate.

5.1 PSYCHOACOUSTIC MODEL OF HEART SOUNDS AND MURMURS

In the early days of medical science as well as today, a stethoscope is a primary medical device used by the doctors for listening the heart sounds and murmurs. Now, the cardiologists use a high sensitive cardiac stethoscope for clinical investigations of the cardiovascular disease. Cardiac auscultation is highly subjective and a cognitive process and the amount of information that can be obtained by listening heart sounds largely depend on the expertise, experience and acuity of the ear of the physician. In general, the classification and interpretation of heart sounds and murmurs is based on a subjective 0-

6/6 grade scale and described by using “faint”, “soft”, “loud”, “ high pitch”, “clear”, “thrill”, “tremor”, “musical” and others terms. These terms are not well-defined and suitable mathematical models are not available. Apart from that, the adjective scales vary among the doctors and difficult to derive a standard model for heart sound quality and correct clinical interpretation. Hence, this thesis work is motivated by the need for a psychoacoustic model of heart sounds and murmurs. A set of set of objective functions, mathematical models are derived and experimentally obtain the various parameters (pitch, intensity, faint, thrill, etc.) that characterize the heart sounds and use them as feature vectors for searching, retrieval and classification applications [McNab et el., 1996), Corwin (2006)]].

The heart sounds and murmurs are an acoustic phenomenon caused by the mechanical events of the heart and thoracic region. The heart sounds and murmurs can be heard by the doctors and cardiologist using conventional acoustic stethoscope. In phonocardiography (PCG), it is also possible to record the heart sounds and murmurs as function of time (time series data) and displayed on a video terminal unit or printed on a thermal paper. The PCG analysis — the art and science of recording and interpreting of heart sounds using latest digital technology (e.g., electronic stethoscope) has significantly helped us to understand and interpret the complex heart sounds (normal, abnormal sounds including murmurs) and in particular valvular diseases. The PCG is a time series data and frequency components (e.g., spectrogram) display of the heart sounds and murmurs can provide useful clinical information to the physician and complements cardiac auscultation

This work describes, psychoacoustic models based on a psychoacoustic principles and mathematical foundations and discussed the psychoacoustic features (pitch, intensity, timbre, loudness, power, intensity and other clinically important psychoacoustic features sharpness, roughness, sound strength, fluctuation strength) that can be modeled, analyzed and provide effective aid of clinical decisions related to heart diseases, and in particular murmurs. These models offer a reasoning framework for the subjective reasoning of heart sounds and derived psychoacoustical models. The classification of heart sounds is characterized by using the psychoacoustic features and derives mathematical equations.

It is also used to model the quality of heart sounds for many standardization efforts and can be used as an effective teaching aid for the cardiac auscultations. The initial investigations and experimental results based on psychoacoustic models are quite encouraging and provide a deeper insight into the perception and interpretation of cardiac auscultations.

Psychoacoustics is the study of the subjective human perception of sounds [E. Zwicker and H. Fastl (1999)]. Alternatively it can be described as the study of psychological correlates of the physical parameters of acoustics. The field of psychoacoustic aims to model parameters of auditory sensation in terms of physical signal parameters and provide a framework and modeling capabilities for the acoustic sounds. The psychoacoustic models of sound perception exploiting the imperceptible sounds are used in the audio compression such as mp3 standards; non-linear response of the ear is exploited in the noise reduction systems and communication networks. The human ear can nominally hear sounds in the range 20 Hz to 20,000 Hz (20 kHz). Frequency resolution of the ear is 0.36 Hz within the octave of 1,000–2,000 Hz. That is, changes in pitch larger than 0.36 Hz can be perceived in a clinical setting. Other scales have been derived directly from experiments on human hearing perception, such as the Mel scale and Bark scale and these are approximately logarithmic in frequency at the high-frequency end, but nearly linear at the low-frequency end.

Ear drums are sensitive only to variations in the sound pressure, but can detect pressure changes as small as 2×10^{-10} ATM and as great or greater than 1 ATM. The sound pressure level (SPL) is also measured logarithmically, with all pressures referenced to 1.97385×10^{-10} ATM. The lower limit of audibility is therefore defined as 0 dB, but the upper limit is not as clearly defined. By measuring this minimum intensity for testing tones of various frequencies, a frequency dependent absolute threshold of hearing (ATH) curve may be derived. Typically, ear shows a peak of sensitivity (i.e., its lowest ATH) between 1 kHz and 5 kHz, though the threshold changes with age, with older ears showing decreased sensitivity above 2 kHz. Equal-loudness contours indicate the sound pressure level (dB), over the range of audible frequencies, which are perceived as being of equal loudness and may be plotted. We use classical reference [E. Zwicker and H. Fastl (1990)] which represents a set of algorithms for calculating auditory sensations including loudness, sharpness, roughness, softness, sound strength and intensity, and fluctuation strength and extend it for the heart sounds and murmurs. The classification of murmurs is characterized by using the psychoacoustic features and derives mathematical equations.

5.1.1 LOUDESS OF HEART SOUNDS AND MURMURS

The loudness is modeled by the following equation, where N is loudness, N' is the loudness of a given critical band or also know as specific loudness, and dz is the increment in the critical band scale [Sun (2000)].

$$N = \int_{0}^{24} N'dz \tag{5.6}$$

The unit of loudness, the sone, is a ratio scale referenced against the sensation produced by a 1 kHz sine tone with a sound pressure level of 40 dB. The referenced scale here refers to equation 5.2 and 5.6.

$$p_2 = 2 \text{ x } 10^{-5} \text{ N/m}^2 \text{ and } I_2 = 10^{-12} \text{ W/m}^2 \tag{5.7}$$

Several experiments were conducted to derive the loudness of heart sounds and murmurs of various diseases and are tabulated in the Table 5.1. The audio recordings of the heart sounds and murmurs subjected to the audio analysis using various DSP techniques using the JAudio and Matlab tools. The user can select the various features and derive the empirical values of the various disease and murmurs. For example, the systolic mitral regugration has high frequency and high loudness that doctor uses for the clinical interpretations. The test samples consist of the PCG recording of the systolic mitral regugration from the audio heart database and executed using JAudio tool derive a set of spectral and psychoacoustic vector features. Using the same technique, we derived the frequency of various heart disease and murmurs and plotted in the Figure 5.2. It is observed that the frequency components show large variations (e.g., systolic mitral regugration – 900 Hz and diastolic rumble – 8 Hz) and can be used a distinguishing feature for searching and classification algorithms [Ristama (1967)].

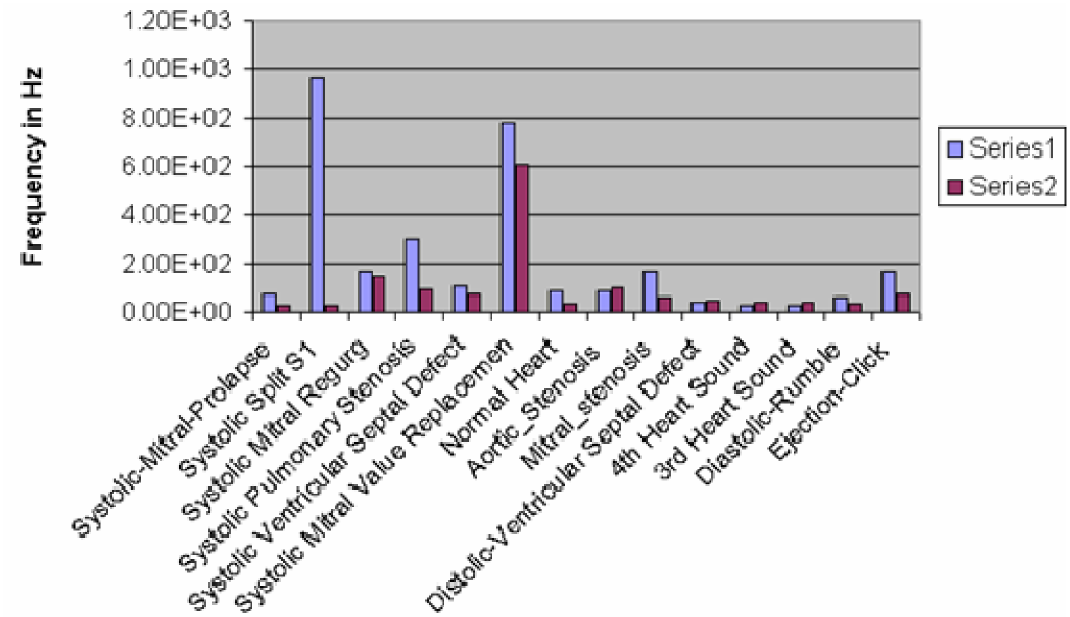


Figure 5.2 Low frequency (Hz) of various diseases.

Based on the frequency component of various PCG signals, we could able to detect abnormalities of different range as shown in the above figure 5.2. Loudness of different PCG is estimated as shown in table 5.1 and its visual representation is shown in figure 5.3. In the table 5.1, the frequencies of various heart diseases were derived by using two methods for better estimation and correct results: (a) spectral centroid (b) Fourier transform

The spectral centroid is main frequency components concentrate and used for the calculation of frequency of the heart sounds and murmurs. Fourier transform is another popular spectral analysis approach for the determination of frequency of the biomedical signals. The average value of both the frequency is used to arrive at correct estimation of frequency of heart sounds and murmurs. Once the frequency of the heart sounds and murmurs is derived correctly, a set of psychoacoustic feature vectors such as loudness or pitch as they are used as discriminatory feature vectors. In the table 5.1, experimentally derived the loudness in dB of various heart sounds and murmurs are described and used for the audio retrieval applications.

Figure 5.4 describes clustering of PCG signals based on loudness property, which will be very useful for diagnosis of abnormalities in heart sounds. The data from the table 5.1 is used for the feature vector – loudness (dB) is used for the clustering the data and visualized in the user interfaces. The clustering of the data on frequency (Hz) and pitch (dB) depicts the clustered data at the low frequency range (< 200 Hz) and more data points in the 30-40 dB loudness range.

Table 5.1 Loudness of different heart sounds and murmurs.

PCG ID	Heart Disease and Murmurs	Frequency Via Spectral Centroid	Frequency Via FFT	Loudness in dB
1	Systolic-Mitral-Prolapse	6.93E+01	2.69E+01	50
2	Systolic Split S1	9.63E+02	2.94E+01	11
3	Systolic Mitral Regurg.	1.62E+02	1.45E+02	27
4	Systolic Pulmonary Stenosis	2.94E+02	9.53E+01	21
5	Systolic Ventricular Septal Defect	1.07E+02	7.55E+01	22
6	Systolic Mitral Value Replacement	7.79E+02	6.09E+02	9
7	Normal Heart	8.89E+01	2.98E+01	41
8	Aortic_Stenosis	8.42E+01	9.70E+01	28
9	Mitral stenosis	1.67E+02	5.66E+01	23
10	Distolic-Ventricular Septal Defect	3.81E+01	4.64E+01	54
11	4th Heart Sound	2.62E+01	3.76E+01	62
12	3rd Heart Sound	2.70E+01	3.86E+01	59
13	Diastolic-Rumble	5.79E+01	3.35E+01	48
14	Ejection-Click	1.67E+02	7.08E+01	24

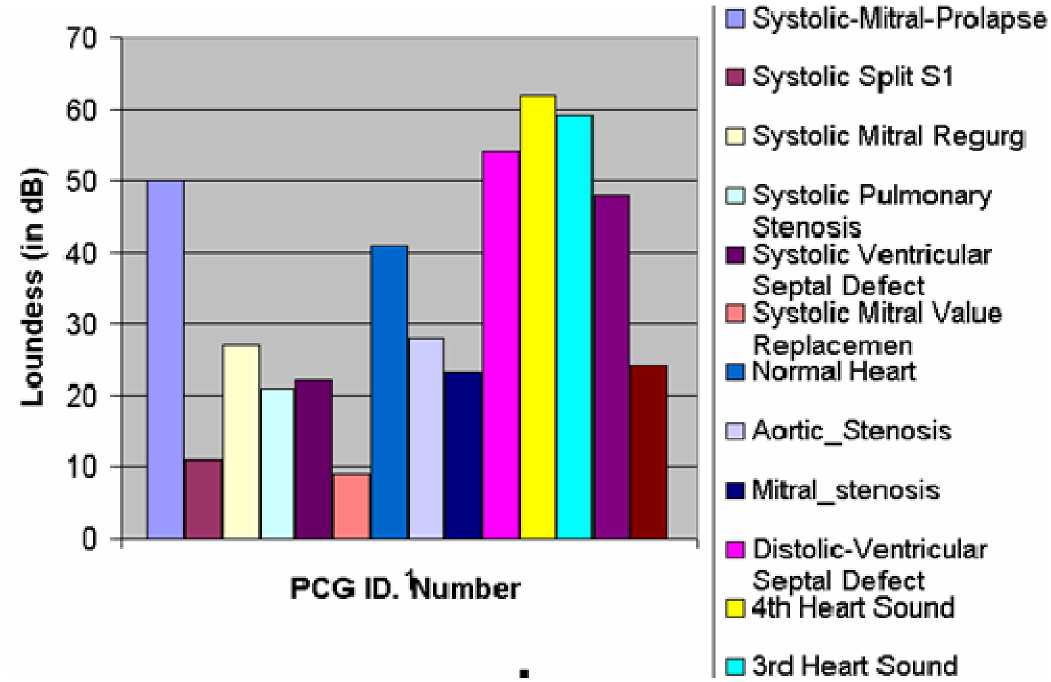


Figure 5.3 Visualization of loudness (in dB) of various diseases.

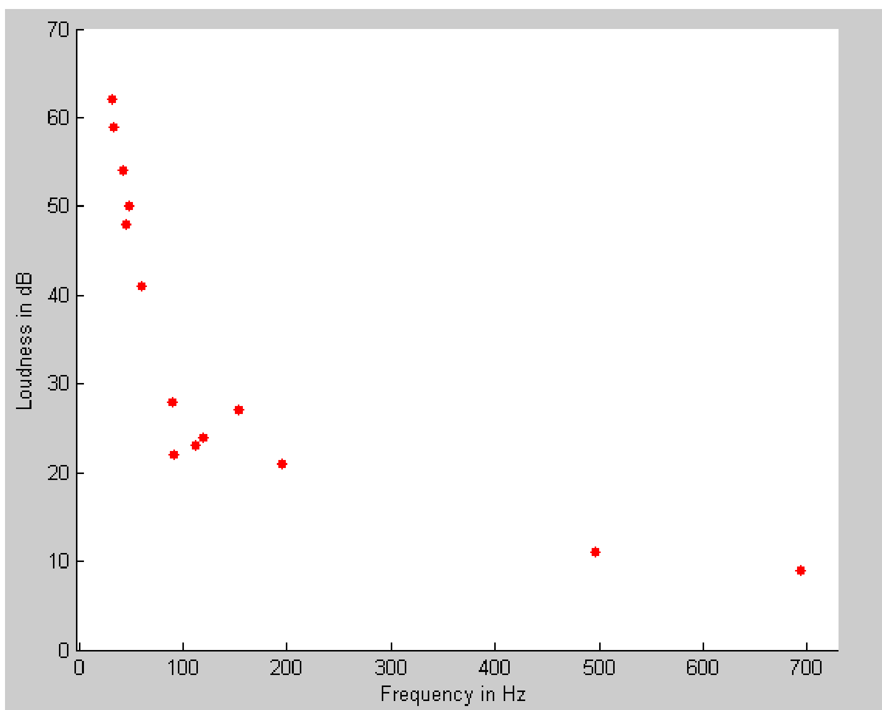


Figure 5.4 Clustering of PCG based on feature vector – loudness.

5.1.2 SHARPNESS OF HEART SOUND AND MURMURS

Sharpness or brightness is one of the most prominent features of timbre. Timbre is more complicated, being determined by the harmonic content of the signal. The hearing is based on the amplitude of the frequencies and is very insensitive to their phases. The shape of hearts sounds and murmurs in time domain waveform is only indirectly related to hearing and poses serious challenges in correct interpretation of heart sounds. The models are based on the centroid (signal spectrum or loudness pattern) of the heart sounds and murmurs. The sharpness is modeled as a weighted centroid of the specific loudness pattern. The unit is acum, referenced to a band of noise 1 critical band wide, centered on 1 kHz at 60 dB. It is also referred to the perception that the sound is "sharp", "harsh" or "soft" when used in the context of heart sound perception and clinical interpretations. It is related to the proportion of high frequency energy present in the sound, weighted towards energy in the region above 3 kHz. For harmonic tones, sharpness can be controlled through distribution of the harmonic spectral envelope. The model proposed in [E. Zwicker and H. Fastl (1990)] is used for the calculating the sharpness of tones is summarized by the Equation 5.8.

$$s = 0.11 \frac{\int_{0}^{24 Bark} N' g(z) z dz}{\int_{0}^{24} N' dz} \qquad (5.8)$$

where the S is sharpness, N' is specific loudness, z is the bark scale of auditory filters and g(z) is a weighting function that emphasis z for the critical band rates.

The timbre or sharpness of the heart sounds and murmurs can be characterized by using the spectral centroid. The spectral centroid is derived by using the frequency analysis using the FFT of the heart sound recordings. The experimental data based on the characterization of sharpness is described in the table 5.2 for various diseases and murmurs. T may observe that the experimental data clearly shows large variations in the sharpness as perceived by the doctor and is also plotted in the Figure 5.5. A systolic split S1 is sharp sound and compares with normal heart sounds. The other heart diseases and murmurs have less sharp audio perception is concerned.

Table 5.2 Estimation of spectral centroid for various diseases.

PCG ID	Disease Name	Spectral Centroid
D1	Systolic-Mitral-Prolapse	2.22E+00
D2	Systolic Split S1	3.08E+01
D3	Systolic Mitral Regurg	5.63E+00
D4	Systolic Pulmonary Stenosis	9.41E+00
D5	Systolic Ventricular Septal Defect	3.42E+00
D6	Systolic Mitral Value Replacement	2.49E+01
D7	Normal Heart	2.85E+00
D8	Aortic_Stenosis	2.70E+00
D9	Mitral stenosis	5.34E+00
D10	Distolic-Ventricular Septal Defect	1.22E+00
D11	4th Heart Sound	8.37E-01
D12	3rd Heart Sound	8.64E-01
D13	Diastolic-Rumble	1.85E+00
D14	Ejection-Click	4.34E+00

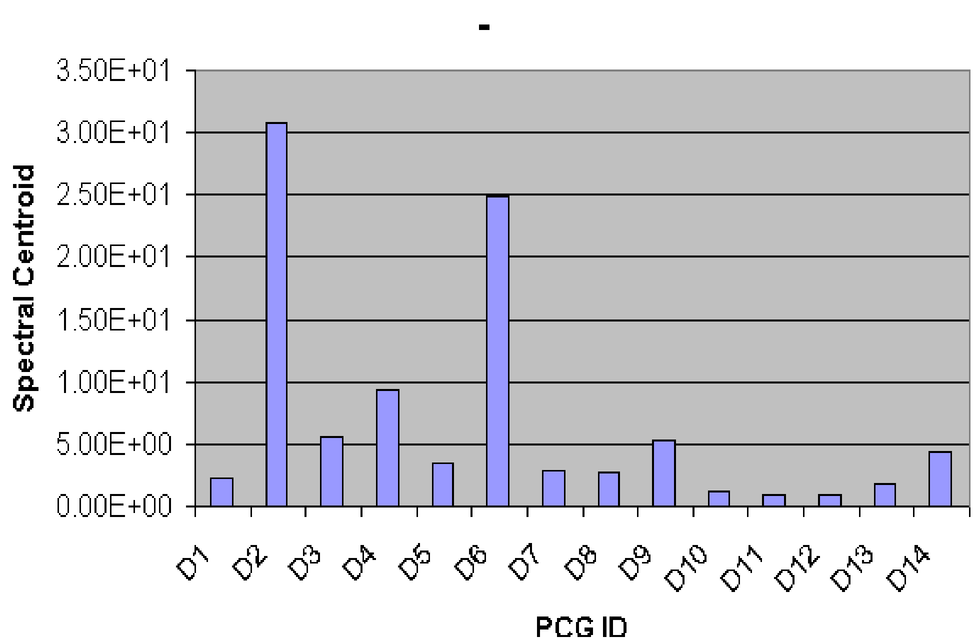
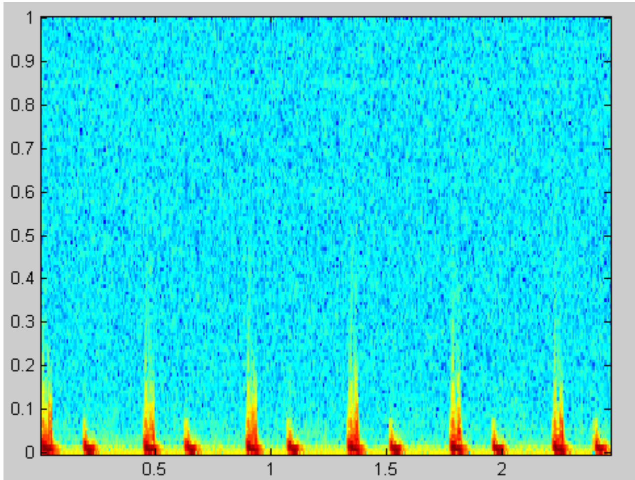
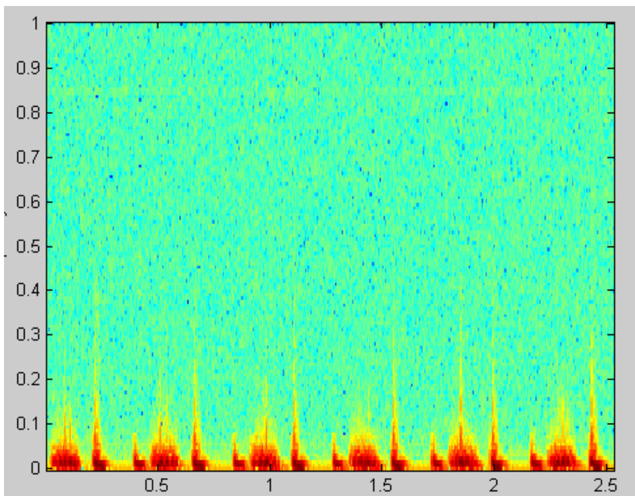


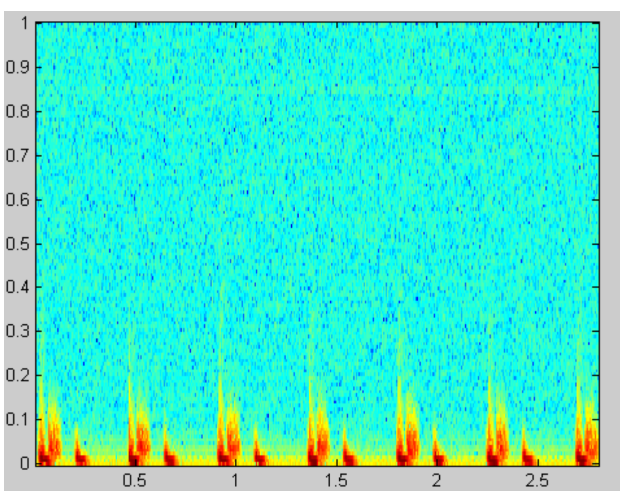
Figure 5.5 Visualization of spectral centroid of various disease and murmurs.



a. Normal heart sound spectrogram



b. Diastolic rumble spectrogram



c. Early systolic murmur spectrogram

Figure 5.6 Sample spectrogram of various diseases (a) normal (b) diastolic rumble (c) systolic murmur.

A spectrogram is a time-varying spectral representation that shows in Figure 5.6 how the spectral density of a signal varies with time. Here also, we can observe large variations in spectrogram representation of PCG signal is helpful for better understanding to signal and visual inspection tool in designing graphical user interference for retrieval of heart sounds and murmurs.

5.1.3 PITCH OF HEART SOUNDS AND MURMURS

A psychoacoustical pitch ratio scale is a difficult concept due to the complexity of pitch perception and cognition [B.C.J More, B.R. Glasberg and T. Baer, 1976] describes some of the complexity of pitch structures for harmonic tones (such as pitch height, octave equivalence and cycle of fifths) through multidimensional geometric figures. In addition to complex structures of pitch height, pitch has the dimension of pitch strength, also known as "tonal ness". The harmonic series is of great importance in pitch perception, and mainly pitched sounds in everyday experience exhibit harmonic spectra. In general, it is usually determined by the fundamental frequency as a pitch percept. A model of pitch perception is analyzed using template matching or autocorrelation techniques [Johnston (1988)].

Pitch is used to describe the tonal quality of the murmur be it high pitched or low pitched. For those of us not musically inclined, a simple way to distinguish pitch is to determine whether the sound is heard best with the diaphragm of the stethoscope, i.e., high pitched, or with the bell, i.e., low pitched. Murmurs of mitral or tricuspid stenosis are best heard with the bell. Some of other hearts sounds that can characterize the pitch are: *S3 sound* is a low frequency, mid diastolic sound occurring 0.14 - 0.22 second after S2 [Dubnowski et el., (1976)]. The frequency components of low frequency heart sounds are difficult to hear and can be modeled and uniquely find pitch features of the heart sounds and murmurs [P.Stein (1981)]. *S4 sound* is also a low frequency, late diastolic sound occurring 0.08 - 0.20 second prior to S1. It is generated during pre-systolic ventricular filling due to atrial contraction, hypertension and diastolic dysfunction. The sequencing and ordering with respect to the S1 and S2 in a standard cardiac cycle and obtain pitch pattern and frequency components using spectral techniques will assist the doctor for the better clinical decisions [P. Stein (1981)].

5.1.4 ROUGHNESS OF HEART SOUNDS AND MURMURS

Roughness is a sensation caused by quite rapid amplitude modulation within auditory filters. The unit of roughness is the asper, which is referenced to a 1 kHz tone at 60 dB with 100% amplitude modulation at 70 Hz. The model presented in [E. Zwicker and H. Fastl(1999)] by for calculating the roughness of modulated tones having a single modulation frequency is given in equation 5.9, where R is roughness, f_{mod} is the modulation frequency, and L_E is the excitation level within an auditory filter. This uses the time varying excitation pattern of the ear (similar to the specific loudness pattern, except that the magnitude is in decibels rather than sones/bark), with the difference between maximum and minimum excitation levels integrated across auditory filters used to determine roughness.

$$R = 0.3 \frac{f_{\text{mod}}}{kHz} \int_{0}^{24Bark} \frac{\Delta L_E(z)dz}{dB/Bark} asper \qquad (5.9)$$

The roughness of the heart sound can be used for the modeling aortic insufficiency (AI) may be congenital rheumatic, and collagen vascular disease. The murmur is a high frequency (blowing) decrescendo murmur beginning in early cardiac cycle and uniquely radiates to the top of the head. The roughness of the murmurs can be uniquely characterized using the above equation and needs further investigations. Mitral regurgitation (MR) is associated with endocarditic and ischemic heart diseases. The murmur is typically a high frequency, holosystolic, plateau murmur that is best heard at the apex. The murmur often radiates to the left axilla and back.

There is no appreciable change in murmur intensity with cycle length (as with AS). MR may be associated with S3 in more severe cases. Here better to use intensity, high frequency and radiation patterns in a consistent way and need further investigation and clinical validations. These models offer a reasoning framework for the subjective reasoning of heart sounds and derived psychoacoustical models. It is also used to model the quality of heart sounds for many standardization efforts and can be used as an effective teaching aid for the cardiac auscultations. The preliminary investigations and experimental results on psychoacoustic models are quite encouraging and provide a deeper insight into the perception and interpretation of cardiac auscultations.

5.2 CONTENT BASED MATCHING AND RETRIEVAL TECHNIQUES

The human ability to distinguish between different PCG signals has been a subject of investigation for a number of years. By definition [Christopher (2005)] that quality of auditory sensation by which a listener can distinguish between two sounds of equal loudness, duration and pitch is known as timbre. Hence it could be said that sound recognition is largely dependent on timbre. Unfortunately, unlike pitch and loudness, timbre has proven to be somewhat difficult to measure or quantify. It is recommended to use discriminatory vectors to distinguish between various PCG signals and detect abnormalities using MFCCs {Logan, ()]. It also explores how many of these coefficients are necessary and useful for accurate classifications of heart diseases.

The use of Mel-frequency Ceptral Coefficients in the identification of PCG (phonocardiography) signal has been examined. Various heart sounds are analyzed to extract their coefficients. These coefficients are reduced using principal component analysis. These coefficients are reduced using principal component analysis. Multi-layered perceptron is trained using principal components. The network is trained for MFCC values of normal heart sound. This trained network is then used to classify different input heart sound samples. By training and testing the network on a different number of coefficients, the optimum number of coefficients to include for identifying abnormality with heart functioning is identified. From the results we can conclude that at least 10 MFCCs should be used. It was observed that taking 3 principal components gave the best classification and that the highest result was obtained from using 13 MFCCs [Kemp, (1995), Logan, (2000)].

Content-based indexing of audio data has become more important since conventional databases cannot provide the necessary efficiency and performance [Foote, (1997)]. However, there are few main difficult problems. First, the content of audio data is subjective information; it is hard to give the descriptions in words. The recognition of data content requires prior knowledge and special techniques in Signal Processing and Pattern Recognition, which usually require long computing time. Second, since several audio features can be used as indices (such as pitch, amplitude, and frequency), a method or processing technique designed and developed for one feature may not be appropriate for another. Thus, there arises the necessity of content based analysis of PCG signal. Here, we discuss a PCG retrieval system, based on modular architecture consisting of four stages. Feature extraction, histogram modeling, pattern matching using different techniques for retrieval and classification.

Histogram modeling using MFCC feature vectors and various techniques of similarity measures are discussed for pattern matching of PCG signals [Keislar et el., (1995)]. Search accuracy of histogram matching is tested with respect to correlation algorithms and it is observed that histogram algorithm for content based heart sound retrieval is more efficient and accurate.

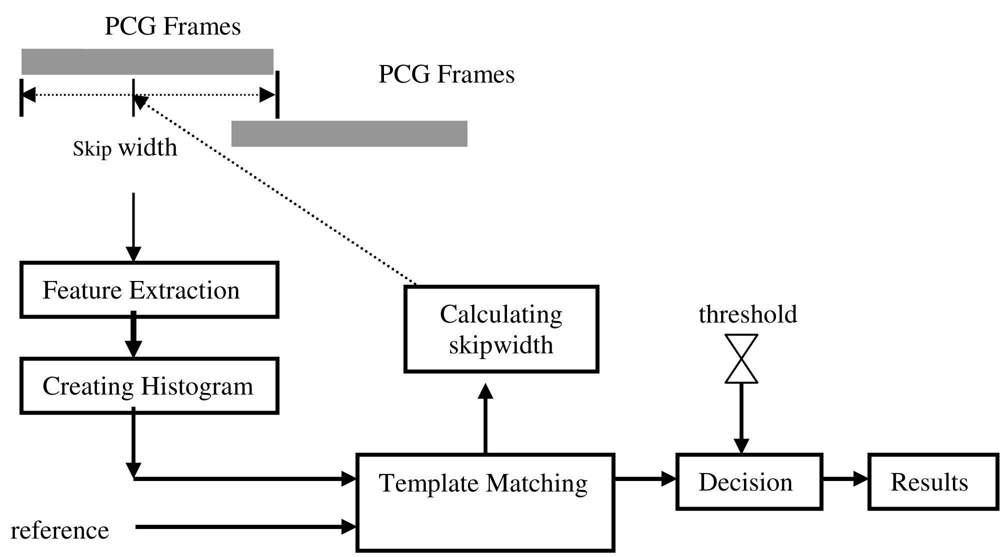


Figure 5.7 Block diagram of content based PCG retrieval algorithm.

The block diagram PCG retrieval algorithm is as shown in Figure 5.7. Firstly, the feature vectors are extracted from the reference PCG and test PCG signal. The window on the input signal is shifted with a range of overlapping.

The window length may be same as the reference audio duration. The feature vectors are used to create histogram. The template matching is carried out to detect and locate the reference PCG from the input PCG.

A PCG feature vector *feature (n)* is written as

$$\text{feature}(n) = (f(n), g(n)) \tag{5.10}$$

$$f(n) = (f_1(n), f_2(n), f_3(n), \ldots f_M(n)) \tag{5.11}$$

$$g(n) = (g_1(n), g_2(n), g_3(n), \ldots g_M(n)) \tag{5.12}$$

where n is the time frame. M denotes the number of frequency sub-bands. An element of f (n) is the normalized short-time power spectrum, which is given as

$$f_i(n) = \alpha(n) \times E_i(n) \tag{5.13}$$

An element of g(n) is normalized short-time power spectrum change ratio by time, which is given as

$$g_i(n) = \beta(n) \times ECR_i(n) \tag{5.14}$$

$$ECR_i(n) = (E_i(n) - E_i(n-1))/E_i(n-1) \tag{5.15}$$

Where $E_I(n)$ denotes the energy of output of the I-th sub-band filter at n-th frame.

Because short-time energy is sensitive to high voltage, this algorithm uses short-time average amplitude to carve the change of signal amplitude, which is given as

$$E_i(n) = \sum_{t=n*L}^{(n+1)*L-1} |x_i(t)| \tag{5.16}$$

Where, L is the length of the frame , $x_i(t)$ denotes the amplitude of ith sub-band at sampling point t. And $\alpha(n)$ *and* $\beta(n)$ is normalized constant defined as

$$\alpha(n) = \frac{1}{\max_i(E_i(n))}$$

$$\beta(n) = \frac{1}{\max_i(ECR_i(n))}$$

5.2.1 FEATURE EXTRACTION

Feature extraction is based on a variant of the Mel-frequency Ceptral coefficient (MFCC) representation. MFCCs are commonly used in speech recognition systems because they provide a concise representation of spectral characteristics. Each coefficient has a value for each frame of the sound. The changes within each coefficient across the range of various PCG signal are examined here. Obtaining the MFCCs involves analyzing and processing the sound according to the following steps [Logan, (2000)].

1. Divide the signal into frames.
2. Get the amplitude spectrum of each frame.
3. Take the log of these spectrums.
4. Convert to the Mel Scale.
5. Apply the Discrete Cosine Transform (DCT).

5.2.2 HISTOGRAM MODELING

After feature extraction, we need to train model for each PCG clip, some modeling approaches, such as GMM, HMM, SVM [Duxbury et el., (2003), Khokhar, (2000)], have already been employed for audio modeling. However, because of computationally expensive processing, it is hard to meet the speed demand of quick audio search and retrieval. Histograms can be used as type f non-parametric signal model for both the reference and input signals over a shifted window. It doesn't need computationally expensive processing while it is relatively stable under adverse environments [E. Zwicker and H. Fastl (1999)]. Thus adopt histogram modeling for specific audio detection. .For the sake of removing the influence of noises, the feature vector, firstly, needs to be quantized (VQ) before modeling histogram. We use the codebook of VQ to build histogram. The similarity distance between the reference and input feature vector histogram can be measured by histogram intersection. The histogram intersection for a window is defined as

$$S_n(h^R, h^T(n)) = \frac{1}{L}\sum_{j=1}^{L} \min(h_j^R, h_j^T(n)) \tag{5.17}$$

where h^R is the histogram for the reference; $h^T{}_j(n)$ *is the* histogram started from the i-th frame; and L denotes the number of bins.

The prediction of similarity upper bound and skip width

As the window for input signal shifts forward in time, the similarity based on reference and input feature vector histograms changes with regard to the correlated overlapping between reference and object segment in input stream. We, thus, may predict the next upper bound of the similarity in terms of current value. The upper bound on S(h^R, $h^T{}_i$) can be defined as:

$$S_{ub}(h_i^R, h_i^T(n2)) = S(h_i^R, h_i^T(n1)) + \frac{n2 - n1}{p_i} \tag{5.18}$$

.

Where $h^T{}_i(n_1)$ and $h^T{}_i(n_2)$ are the histograms for windows started from n_1 and n_2 frames respectively, $n_1<n_2$; P_i denotes the total number of frames in each histogram. When the window is shifted the n_2-th frame, $S_{ub}(h^R{}_i, h^T{}_i(n_2))$.

A set of representative MFCC histogram of various heart diseases and murmurs are shown in Figure 5.8. These MFCC histograms are used extensively in the experimental work and correlation factor was measured. The best histogram match will have highest correlation and is used the correlation factor as index for searching and retrieval applications.

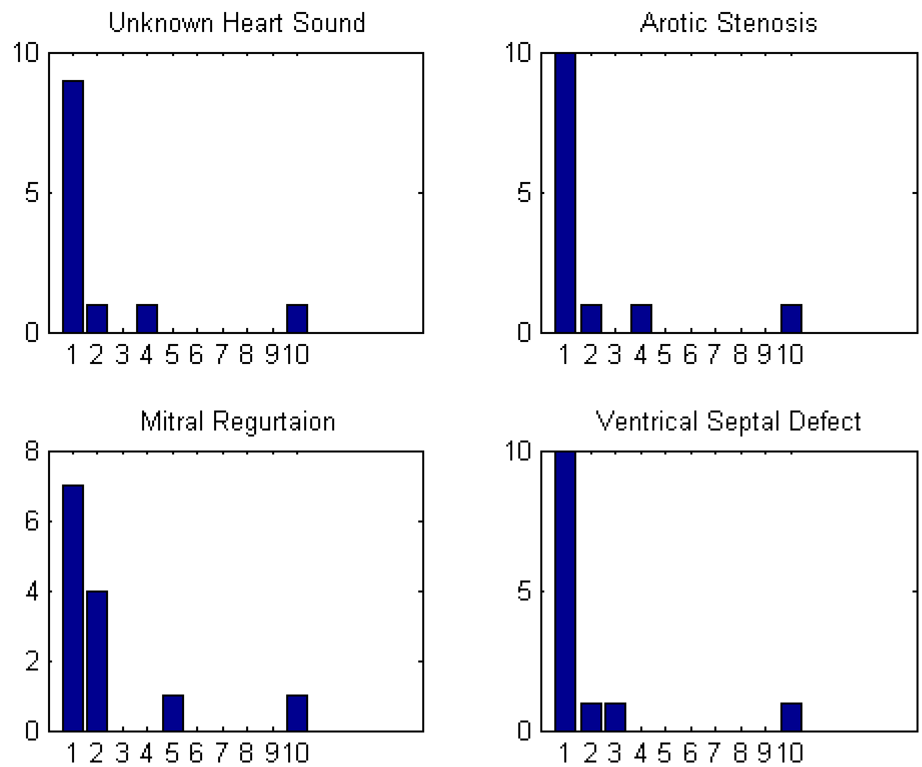


Figure 5.8 MFCC based histogram representation of heart diseases and murmurs.

5.3 OTHER TIMBERAL FEATURES

Spectral Centroid is the centroid of the magnitude spectrum of short-term Fourier transform and is a measure of spectral brightness. *Spectral Rolloff* is the frequency below which 85% of magnitude is concentrated. It measures the spectral shape. *Spectral Flux* is the squared difference between the normalized magnitudes of successive spectral distributions. It measures the amount of local spectral change. *Zero Crossing Rate* (ZCR) is the number of time domain zero crossings the signal. The ZCR is used for the characterization of noise-like features as used a discriminatory feature for heart sound and murmurs database searching and retrieval applications [Sun, (2002)].

5.3.1 CALCULATING PCG SIMILARITY

Three different measures are available for computing the similarity between chord-histograms. Cosine Similarity, Chi-Squared Similarity and Euclidean distance (with normalization), given in equations (5.19), (5.20), and (5.21) respectively. Cosine similarity and Euclidean similarity are used to calculate. In each of these equations, *A* and *B* represent the *k*-dimensional histogram vectors $\varphi(a)$ and $\varphi(b)$ of audio recording of heart sounds *a* and *b*.

$$sim_{ab} = \cos^{-1}\left(\frac{A \cdot B}{||A||\,||B||}\right) \tag{5.19}$$

$$sim_{ab} = \frac{1}{2}\sum_{i=1}^{k}\frac{(A_i - B_i)^2}{A_i + B_i} \tag{5.20}$$

$$sim_{ab} = \sqrt{\sum_{i=1}^{k}\left(\frac{A_i}{|A|} - \frac{B_i}{|B|}\right)^2} \tag{5.21}$$

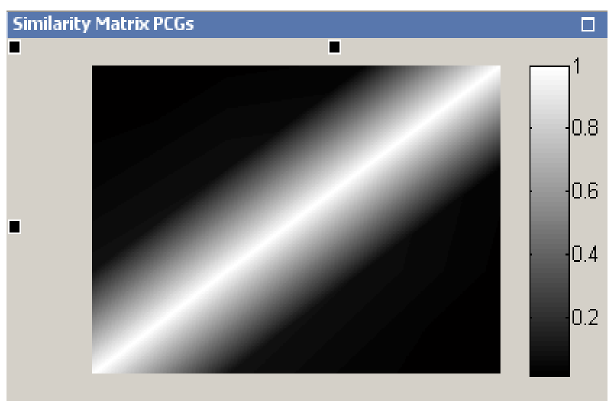


Figure 5.9 Similarity matrix for PCG signals.

For each of these similarity measures, a smaller value indicates a better song match. Similarity matrix of different PCG signal is as shown in Figure 5.9.

5.3.2 SEARCH ACCURACY

The search accuracy was evaluated by the recall rate δ, precision rate ξ, and η average accuracy. These are defined as

$$\delta = \frac{\text{the number of correctly retrieved objects}}{\text{the number of objects should have retrieved}} \tag{5.22}$$

$$\xi = \frac{\text{the number of correctly retrieved}}{\text{the number of all retrieved objects}} \tag{5.23}$$

$$\eta = (\delta + \xi) / 2 \tag{5.24}$$

Recall is typically used in conjunction with precision, which measures the fraction of the retrieved patterns that is relevant. Precision and recall can often be traded off, i.e. one can achieve high precision and low recall or the other way round.

5.4 EXPERIMENTS AND RESULTS

A series of experiments were conducted to know the precision and retrieval efficiency. About 200 heart sounds and murmurs covering 30 types of cardiovascular pathologies were stored in the database by acquiring from the electronic stethoscope. Some of the heart sounds and murmurs were collected from the Texas Heart Research Center, USA. The heart sounds and murmur database contained the normal and benign heart sounds and innocent murmurs. Abnormal heart sounds and murmurs were compared with the normal sounds using similarity measures using MFCC. The MFCC shows remarkable distinguishing features in normal heart sound and murmurs. When comparing with murmurs, the ZCR plays critical role in separating murmurs and number of the sign change determines the murmurs vector features.

The retrieval of S3 and S4 hearts sounds was difficult and performed poorly by retrieving 21 false audio files in a collection of 10 relevant files for S3 and S4. The characterizations of S3 and S4 features vectors in terms of pitches was inadequate and recommend using other features such as magnitude spectral or spectral centroid. In our experiments, the similarity measures for detecting abnormality is calculated to be 100% and experiments also records a detection of murmurs (early systolic and diastolic rumble) showed a better precision of 80% match to the unknown murmur. The ranked indexing with similarity measure is difficult to model using only feature vector. For example, if we use pitch as indexing parameter and many heart sounds and murmurs file retrieved and inadequate for indexing and retrievals. The performance of the histogram search in comparison with correlation coefficient matching showed that the histogram search algorithm can achieve a high precision about 97%.

5.5 SUMMARY

The cardiac auscultation is an effective diagnostic technique used in the early detection of cardiac diseases in particular the valvular diseases, including the murmurs. It is argued that the most of the doctors and physicians depend on their experience and make subjective interpretations of heart diseases. A psychoacoustic models based on a psychoacoustic principles and mathematical foundations and discussed the psychoacoustic features (pitch, intensity, timbre, loudness, power, intensity and other clinically important psychoacoustic features) that can be modeled, analyzed and provide effective aid of clinical decisions related to heart diseases, and in particular murmurs. These models offer a reasoning framework for the subjective reasoning of heart sounds and derived psycho acoustical models. It is also used to model the quality of heart sounds for many standardization efforts and can be used as an effective teaching aid for the cardiac auscultations. The investigations and experimental results on psychoacoustic models are quite encouraging and provide a deeper insight into the perception and interpretation of cardiac auscultations.

Content based retrieval of heart sounds is discussed in detail. Histogram modeling using MFCC feature vectors and various techniques of similarity measures are discussed for pattern matching of PCG signals. Search accuracy of histogram matching is tested with respect to correlation algorithms and it is observed that histogram algorithm for content based heart sound retrieval is more efficient and accurate.

CHAPTER – 6

SEARCHING RETRIEVAL AND CLASSFICIATION OF HEART SOUNDS AND MURMURS

6.0 INTRODUCTION

In the domain of audio information retrieval applications, one of the challenges is to design of fast searching methods and spatial techniques that will search a database of multimedia objects to locate the objects that match the query object, exactly or approximately [Volker G and Günter. O (1998),]. Objects can be two-dimensional color images, grey scale medical images in 2D or 3D (e.g., MRI brain scans), one-dimensional time-series (e.g., physiological signals), digitized voice or music, video clips etc. Specific applications include image database, financial and product time series; scientific databases with vector fields; audio and video data bases; DNA databases etc. In such database, typical queries would *be "find the medical image X-rays that contain that has the texture of a tumor?"* Searching for similar patterns in such databases as the above is essential, because it helps in predictions, computer aided medical diagnosis, data mining and rule discovery [A. Guttmann, R-trees (1984), Beckmann et el., (1990)].

In this research work, focus is on the time – series audio data which is time-sampled data using digital sampling techniques of the heart sounds and murmurs. It is possible to model the audio objects using the lower level features such as amplitude, frequency etc. and more amenable to the digital processing techniques such the Fast Fourier transforms [Celma and Herrera, (2005]. Such features are called frame level features with or without overlapping. However, it is better to model the audio objects at a higher level based on the lower features. As doctors are acustomized to the hearing of the heart sounds and murmurs at the perception level and may not be interested in the spectral properties or complex digital processing techniques.

A typical clinical investigation reasoning of a doctor or cardiologists is characterized by the perceptual properties such as loudness, pitch intensity, thrill, gallop etc as cues for the clinical diagnosis. From the auscultation practice and medical training the doctor is interested in the query which is based on the audio content and perceptual properties. Typical query may be "*Give all the heart sounds which has loudness grater than 15 dB and have noise-like sound?*"

Here, a formal framework for the indexing and retrieval of heart sounds from the heart databases is discussed. The distance between of two objects, O_1 and O_2 has to be quantified and derive a distance function $\mathcal{D}(\ )$ from the mathematical analysis or characterized by a domain expert.

Definition: Given two objects O_1 and O_2, the distance (= dissimilarity) of the two objects is denoted by

$$\mathcal{D}(O_1, O_2) \qquad (6.1)$$

For example, if the two objects are (equal length) time series data, the distance

$\mathcal{D}(\ )$ could be their Euclidean distance (the root of the sum of the squared differences).

Similarity queries are of two types and can be classified into two categories:

Whole match: Given a collection of N objects $O_1, O_2, O_3 \ldots O_N$ and a query object Q, and the task is to find those data objects that are within distance ε from Q. Notice that the query and the objects are of the same data type i.e., audio.

Sub-pattern Match: Here query is allowed to specify only part of the object. Specifically, given N objects (e.g., images) $O_1, O_2, O_3, \ldots\ldots O_N$, a query (sub-) object and a tolerance ε, we want to identify the parts of the data objects that match the query.

Additional types of queries include the "*nearest neighbors*" queries (e.g., *"find the most similar diastolic murmur sounds to the systolic murmur?"*) and the "*all pairs*" queries or "*spatial joins*" (e.g., *"report all pairs of systolic murmurs that are within distance ε from each other"*. In general, the problem of searching for two dimensional points, which will be organized in R – trees: in this case, nearest - neighbor search can be handled with branch-and-bound algorithms [A. Guttmann, (1984)].

For all above types of queries, the ideal method should fulfill the following requirements.

1. Fast response time within the acceptable tolerance for the searching and distance calculations with each and every object will be too slow for large databases.
2. The results for the query should have be "*correct*" and should return all the qualifying objects, without missing any (i.e., "*false dismissals*"). Notice that false alarms are acceptable and eliminated in the easily through post-processing step.
3. The methods and algorithms should require a small space overhead and space complexity.
4. The method should be dynamic and easy to insert, delete and update objects.

The gist of the presented "approach" is to use f feature extraction functions to map objects into points in f-dimensional space; thus we can use highly fine-tune database spatial access methods to accelerate the search.

6.1 SPATIAL ACCESS METHODS

As mentioned earlier, the idea is to map objects into points in f-D space, and to use multi attribute access method called spatial access methods (SAMs) to cluster them and search for them. Thus, a brief introduction to multidimensional indexing methods is in order. The prevailing methods from three classes are: R* trees and family, linear quad trees and grid-file structure. Several of these methods [A. Guttmann, (1984)] and [N. Roussopoulos (1995)] explode exponentially with the dimensionality, eventually reducing to sequential scanning. For linear quad trees, the effort is proportional to the hyper surface of the query region; hyper surface grows exponentially with the dimensionality. Grid files face similar problems, since they require a directory that grows exponentially with the dimensionality.

The R-tree based methods seem to be most robust for higher dimensions, provided that the fan out of the R-tree nodes remains > 2 [Guttmann, (1984)]. Below, a brief description of the R-tree method and its variants is given, since it is one of the typical representatives of spatial access methods. The R-tree represents a spatial object by its Minimum Bounding Rectangle (MBR). Data rectangles are grouped to form parent nodes, which are recursively grouped, to form grandparent nodes. The MBR of a parent node completely contains the MBRs of its children; MBRs are allowed to overlap. Nodes of the tree correspond to disk pages. Disk pages or disk blocks are consecutive byte positions on the surface of the disk that are typically fetched with one disk access. The goal of the insertion, split, and deletion routines is to give trees that will have good clustering, with few, tight parent MBRs.

A range query specifies a region of interest, requiring all the data regions that intersect it. To answer this query, we first retrieve a superset of the qualifying data region. Compute the MBR of the query region, and then recursively descend the R-tree, excluding the branches whose MBRs do not intersect the query MBR. Thus, the R-tree will give us quickly the data regions whose MBR intersects the MBR of the query region. The retrieved data regions will further examined for the intersection with the query region. Algorithms for additional operations (nearest neighbor queries, spatial joints, insertions and deletions) are more complicated and are still under research [Chen and Chang, (2000)].

6.2 A GENERIC AUDIO INDEXING APPROACH

To illustrate the basic ideas, the problem is formulated as follows and manly focus on the audio object represented by the features and confine "whole – match" queries. For such queries, the problem is defined as follows:

1. There is a collection of N objects $O_1, O_2, O_3, \ldots, O_N$
2. The distance / dissimilarity between two objects (O_1, O_2) is given by the function $D(O_1, O_2)$, which can be implemented as a program
3. The user specifies a query object Q, and a tolerance ε.

The main purpose is to find the audio objects in the heart sounds and murmurs database collections that are within ε from the query object. An obvious solution is to apply a sequential scanning; for each and every object O_i $(1 < i < N)$, compute its distance from Q and report the objects with the distance $D(Q_i, Q_j) < \varepsilon$

However, the sequential scanning may be slow, for two reasons:

1. The distance computation might be expensive. For example, the editing distance in DNA strings requires a dynamic programming algorithm, which grows like the product of the string lengths – typically, in the hundreds or thousands for DNA database.
2. The database size N might be large.

Thus, it is necessary to explore for a faster alternative and can be improved significantly based on two ideas

1. A quick-and-dirty test to discard quickly the vast majority of non-qualifying objects (possibly, allowing some false alarms)
2. The use of spatial methods, to achieve faster than-sequential searching.

Assume that the distance function between two such series S and Q is the Euclidian distance

$$D(S, Q) = \Big(\sum_{i=0} (S[i] - Q[i])^2\Big)^{1/2} \qquad (6.2)$$

where S[i] stands for the value of the observation variable S on the i-th day. Cleary, computing the distance of two stocks will take 365 subtractions and 365 squarings. The idea behind the quick-and-dirty test is to characterize a sequence with a single number, which will help us discard many non-qualifying sequences. Such a number could be, e.g., the average stock price over the year. Clearly, if two stocks differ in their averages by a large margin, it is impossible that they will be similar. The converse is not true, which is exactly the reason we may have false alarms.

The numbers that contain some information about a sequence (or a multimedia object, in general) will be referred to as “features” used for retrieval and classification.. Using a good feature (like the “average” in the stock prices example), we can have a quick test, which will discard many stocks, with single numerical comparisons for each sequence (big gain over 365 subtractions and squarings that the original distance function requires). If using one feature is good, using two or more features might be even better, because they may reduce the number of false alarms (at the cost of making the quick-and-dirty test a bit more elaborate and expensive). Some additional feature might be e.g., the standard deviation or even better some of the Discrete Fourier transforms (DFT) coefficients. The end result of using f features for each of audio objects is to map object into a point in f-dimensional space. It may be referred to as mapping of F ( ) (for ‘F’Feature).

Definition: Let F ( ) be the mapping of objects to f-dimensional points, that is, F (O) will be the f-D point that corresponds O.

This mapping provides the key to improve on the second drawback of sequential scanning: by organizing these f-D points into a spatial access method, it is possible cluster them in a hierarchal structure like the R*-trees. Upon a query, it is possible to explore the R*-tree, to prune out large portions of the database that are not promising. Consider the whole-match query that requires all the objects that are similar to S_n within tolerance ε: this query becomes f-D sphere in feature space, centered on the image F (S_n) of S_n. Such queries on multi-dimensional points are exactly what R-trees and other SAMs are designed to answer efficiently.

The feature vectors play significantly role in indexing and retrieval of audio object - heart sounds and murmurs. Several experiments were conducted and derived a list of significant features that have more discriminatory. For example, zero crossing ratio (ZCR) is a very significant to characterize “noise-like” features and clearly evident in the murmurs. For example, psychoacoustic feature (e.g., loudness) may be used to improve the quality of the search in terms of precision and recall.

The audio search algorithm for a whole match query is described as follows:

6.2.1 SEARCH ALGORITHM

1. Map the query object Q into a point F (Q) in feature space.
2. Using a spatial access methods, retrieve all points within the desired tolerance E from F (Q)
3. Retrieve the corresponding objects, compute the actual distance from Q and discard the false alarms

Intuitively, the method has the potential to relieve both the problems of the sequential scan, presumably resulting in much faster searches. The only step that we have to be careful with is that the mapping F ( ) from objects to f-D points does not distort the distances.

Let D ( ) be the distance function of two audio objects, and D feature ( ) be the (say Euclidean) distance of the corresponding feature vectors. Ideally, the mapping should preserve the distances exactly, in which case the SAM will have neither false alarms nor false dismissals. However, requiring perfect distance preservation might difficult. For example, it is not obvious which features we have to use match the editing distance between two DNA strings. Even if the features are obvious, there might be practical problems and although in theory a SAM can support an arbitrary number of dimensions, in practice they all suffer from the 'dimensionality" as discussed earlier. The crucial observations are that it can guarantee that there will be no false dismissals if the feature space matches or underestimates the distance between two objects. Intuitively, this means that our mapping F( ) from objects to points should make things look closer (i.e., it should be contractive mapping). Mathematically, let O_1 and O_2 be two objects (e.g., same-length sequences) with distance function D ( ) (e.g., the Euclidian distance) and F (O_1) and F (O_2) be their feature vectors (e.g., their first few coefficient), with distance function S feature ( ) (e.g., the Euclidian distance). Then lower bounding lemma is described as follows.

LEMMA 6.1 (LOWER BOUNDING): To guarantee no false dismissals for whole-match queries, the feature extraction function *F* ( ) should satisfy the following formula:

$$D_{\text{feature}}\ (F\ (O_1), F\ (O_2)) \leq D\ (O_1, O_2) \tag{6.3}$$

As discussed in [Y. Manolopoulos, C Faloutsos, N. Ranganathan (1994)], the lower-bounding the distance works correctly for range queries and works for “all pairs” and “nearest neighbor”. An “all pairs” query can easily be handled by a “spatial joint” on the points of the feature space: using a similar reasoning as before, we see that the resulting set of pairs will be a superset of the qualifying pairs. For the nearest neighbor query, the following algorithm guarantees no false dismissals

1. Find the point *F (P)* that is the nearest neighbor to the query point *F(Q)*

2. Put a range query, with query object *Q* and radius e = D (Q, P) (i.e., the actual distance between query object *Q* and data object *P*)

6.2.2 ALGORITHM FOR GENERIC AUDIO OBJECT INDEXING

1. Determine the distance function D ( ) between two audio objects
2. Find one or more numerical feature-extraction functions to provide a “quick-and-dirty” test.
3. Show that the distance in feature space lower-bounds the actual distance D ( ) to guarantee correctness
4. Use a SAM (e.g., an R-Tree) to store and retrieve the f-D feature vectors.

The first step involves domain expert such as doctors or cardiologists posses the medal knowledge about the heart diseases enriched with auscultation skills and can perform a quick clinical decisions. The methodology focuses on the speed of search only; the quality of the results is completely relaying on the distance function that expert will provide. Thus, the algorithm return exactly the same response set (and therefore, the same quality of output, in terms of precision-recall) that would be returned by a sequential scanning of the database and is faster. The second step requires clinical experience and m should focus on the discriminatory features. (referred to as the "feature-extracting" question). Typical feature-extracting question: if we are allowed to use only one numerical feature to describe each data object, what should this feature be? The successful answers to the above question should meet two goals: first, they should facilitate step 3 (the distance lower-bounding) and second, it should capture most of the characteristics of the objects. It may be observed that the mechanism of the quick and dirty filter, in conjunction with lower-bounding lemma, can lead to the two problems:

1. the dimensionality curse (time series)

2. the "cross-talk" or noise of features

For each case study, it is better to describe the audio objects and distance functions, then show how to apply the lower-bounding lemma and finally give experimental results on real or realistic data of the heart sounds and murmurs.

6.3 ONE DIMENSIONAL TIME SERIES

The goal is to reach a collection of equal length time series, to find the ones that are similar to a distance function. The audio objects are considered as one – dimensional and are the sampled data of the heart sounds and murmurs. From the algorithm 6.2, the distance function, first step, is to determine the distance measure between two time series [Y. Manolopoulos C Faloutsos, N. Ranganathan (1994)]. A typical distance function is the Euclidean distance, which is used routinely in financial and forecasting applications and recommended use the same concept to derive the distance functions for the heart sounds and murmurs.

6.4 FEATURE – EXTRACTION AND LOWER BOUNDING

Having decided on the Euclidean distance as the dissimilarity measure, the next step is to find some features that can lower- bound it. A set of features that first, preserve/lower-bound the distance, and second, carry much information about the corresponding time series (so that the false alarms are few). The second requirement suggests the use of "good features" that possess improved discriminatory power. In the murmur searching example, the ZCR is a unique feature that discriminates murmurs and other heart sounds. However, although this would perfectly match the actual distance, it would lead to the dimensionality-curse problem. Cleary, some better features are needed. Applying the second step of the algorithm 6.2, the user can ask the feature-extracting question: *"If we are allowed to use only one feature from each sequence, what would be the feature be?* A natural answer is the average. By the same token, additional features could be the average of the first half, of the second half, of the first quarter, etc. It is recommended to use the coefficients of the Fourier transform or the Discrete Fourier Transform (DFT). For a signal $x = | x_i |$, $i = 0, 1, 2, \ldots, n-1$). The third step of the above methodology is to show that the distance in feature space lower-bounds the actual distance. The solution is provided by Parseval's theorem, which states that the DFT preserves the energy of a signal as well as the distances between two signals

$$D(\vec{x}, \overrightarrow{y}) = D(\overrightarrow{X}, \overrightarrow{Y}) \qquad (6.4)$$

where X and Y are Fourier transforms of x and y respectively. Thus, if the first f ($f < n$) coefficients of the DFT are retained as the features, actual distance measure is lower bound.

$$\begin{aligned}\mathcal{D}_{feature}(\mathcal{F}(\vec{x}), \mathcal{F}(\vec{y})) &= \sum_{F=0}^{f-1} |X_F - Y_F|^2 \\ &\leq \sum_{F=0}^{n-1} |X_F - Y_F|^2 \\ &= \sum_{i=0}^{n-1} |x_i - y_i|^2 \end{aligned} \tag{6.5}$$

And finally

$$\mathcal{D}_{feature}(\mathcal{F}(\vec{x}), \mathcal{F}(\vec{y})) \leq \mathcal{D}(\vec{x}, \vec{y}) \tag{6.6}$$

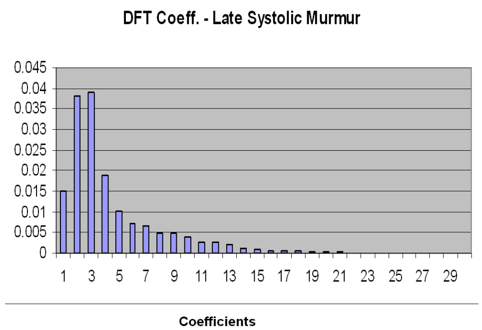
Notice that the discussed approach can be applied to any orthonormal transform such as Discrete Cosine Transform (DFT), wavelet transforms etc., because they all preserve the distance between the original and the transformed space. In fact, response time will improve with the ability of the transform to concentrate the energy: the fewer the coefficients that contain most of the energy, the more accurate our estimate for the actual distance, the fewer the false transform, and faster our response time.

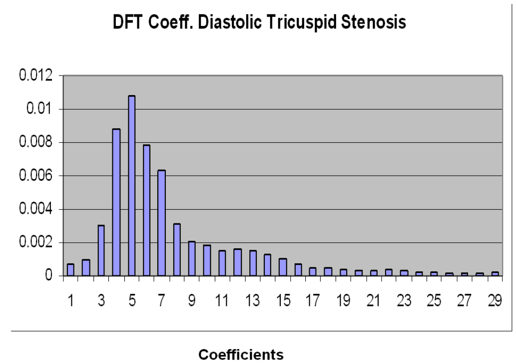
The performance results are encouraging with overheads, but these transforms will achieve even better response times. In addition to being readily avaliable (e.g., in Matlab), it is observed that the DFT coefficients concentrates energy in the first few coefficients, for large class of signals. In the experimental work, the DFT coefficients of various disease is shown in Figure 6.1 and the energy content are mainly confined to the first 30 coefficients and other are very small and can be ignored to achieve better computational efficiency. These signals have skewed energy spectrum (O (F^{-b})) as follows:

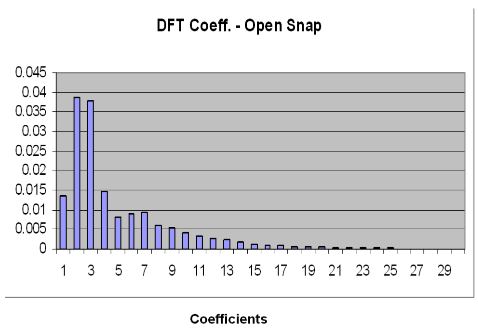
1. For b = 2, results in the so called random walks or brown noise, which model successfully stock movements and exchange rates.

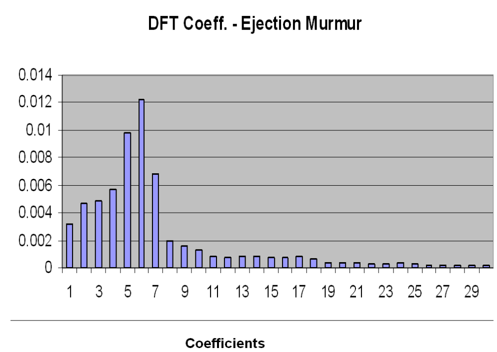
2. With even more skewed spectrum (b > 2), black noises occur. Such signals model successfully, the water level of rivers that rainfall patterns as they vary over time.

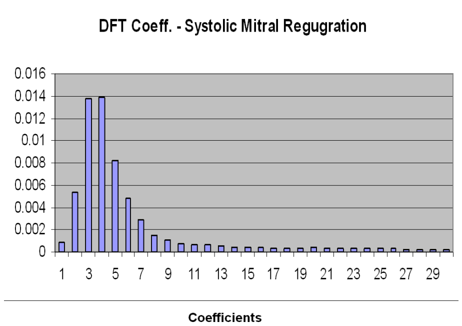
3. With b = 1, we have the pink noise. Birkhoff's theory claims that "interesting" signals such as musical scores and other works of art, consists of pink noise, whose energy spectrum follows O (F^{-1}). The argument of the theory is that white noise with O (F^{0}) energy spectrum is completely unpredictable, while brown noise with O (F^{-2}) energy spectrum is predictable.

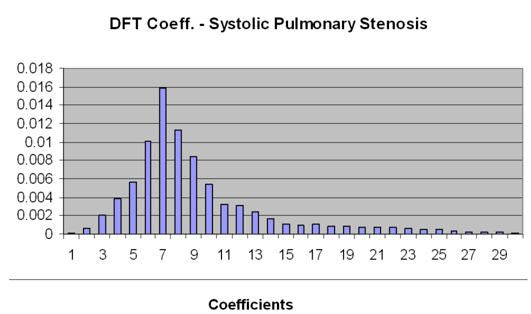


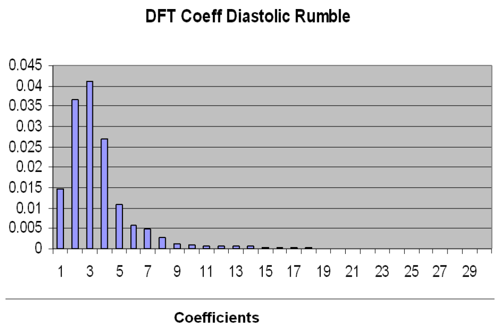












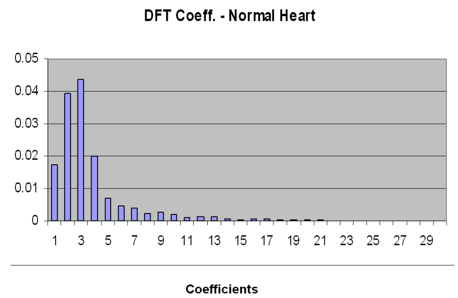


Figure 6.1 The DFT coefficients of various heart diseases and murmurs.

6.5 INDEXING AND RETRIEVAL EXPERIMENTS

Indexing and performance experiments consist of data sets of about the 30 heart disease and murmurs. The heart sound audio collections were colleted from the reference data bases (Texas Heart Center database) as well as our own recordings. The total dataset size is about 500 audio objects (i.e., N = 500). The duration of the each heart sound recordings varies from 1 minute to 5 minutes and 16 bits, mono or stereo recording at various sampling rates (e.g., 16 KHz/Sec). Various types of the heart sound recordings for the same diseases; for example, in ejection murmurs recordings and help us to discriminate between inter-clusters as well intra-cluster classifications and indexing of the audio objects. These recordings are recorded in a clinical setting with minimal background noise. The collection contained various age groups, gender and patient data and is not used for the retrieval purposes; however, it can be used for correlations studies linking age with cardiac disease.

The indexing and retrieval algorithms are implemented in Matlab (version 7.2) and Java (version 1.2) with MySQL databases (version 5.6) and other related utilities. The algorithms were run on a Pentium IV dual core with 50 MHz, 1 GB RAM and 64 GB hard disk. The experimental times were calculated with the software instrumentation API that helps us to find the “elapsed” time between to events and were integrated with the software and tabulated the results. It is better to define the following execution timings so that the results can be interpreted correctly. All the timings are measured in seconds and able to model the performance experiments.

1. Total execution time (Total $_{exec}$) in seconds is the sum of the indexing time and the retrieval time.
2. The indexing time (Time $_{index}$) is the time used to create the indexes in the databases.
3. The retrieval time (Time $_{retrieval}$) is the execution of the distance-based or content based measurement of similarity/dissimilarity retrieval algorithms.

The standard R* - tree was used for the spatial access methods for the experiments. The heart sounds recordings sequences were retrieved form the hearts sound database and the number N varied from 5 to 500 with length n = 1024.

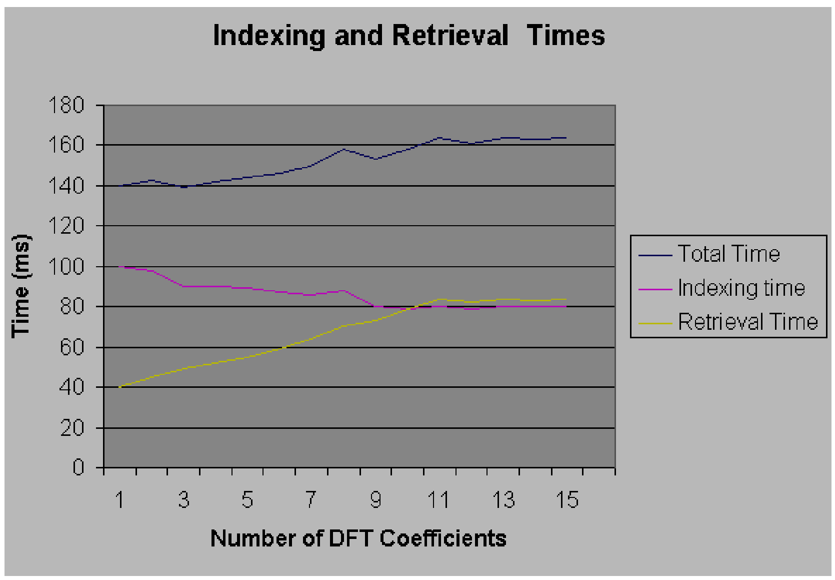


Figure 6.2 The indexing and retrieval times as a function of DFT coefficients.

The above figure 6.2 shows the break-up of the response time as function of the number of DFT coefficients. The diamonds, triangles, and squares indicate total time, post-processing time and R*-tree time, respectively. Notice that, as we keep more features f, the R*-tree becomes bigger and slower, but more accurate (fewer false alarms, and therefore shorter post-processing time). This tradeoff reaches an equilibrium for f = 12 or 13. For the rest of the experiments, the f= 2 Fourier coefficients were kept for indexing, resulting in a four-dimensional R*-tree (two real numbers for each complex DFT coefficients) and response time is optimal but the quality the data results is poor. For various experimentations, we observed that the number of DFT coefficient (12 or 13) is the best with good response time at the same the quality of the results in terms of precision is highly acceptable.

6.6 SUMMARY

Important observations and conclusions may be summarized as:

1. The indexing and retrieval algorithms are well suited for the audio objects – heart sounds and murmurs and can be applied for the indexing, searching and matching applications.

2. The distance measurement algorithms (e.g., Euclidian distance) is computationally simple and varies a as function of the square root or sum squared and has a significant effect when the number of audio objects is large. However, use optimization techniques such as caching and intelligent pre-processing steps can be used.

3. The indexing time is decreased exponentially for small indexing terms and may increase exponential after a threshold value of high degree of dimensionality. ($c = 20$) or as a function of features vectors and their dimensionality.

4. For signals with skewed spectrum (Figure 6.2) a minimum in the response time is achieved for a small number of Fourier coefficients ($f = 10, 12, 13$). The minimum is rather flat, which implies that a suboptimal choice for 'f' will give search time that is close to the minimum and exploit the energy concentrating properties can be exploited.

5. The distance based histogram matching algorithm was also experimented with MFCC coefficients with dimensionality ($c = 12$) and matching algorithms gave a better results (few false alarms) and best retrieval times < 100 ms.

CHAPTER – 7

HEART SOUNDS AND MURMUR SIGNAL FEATURE EXTRACTION AND CLASSIFIER

7.0 INTRODUCTION

In PCG systems, it is required to detect of different murmur sounds and an approach for multiclass signal classification based on second-order statistical feature is proposed in this research. The proposed system is designed to recognize different heart sounds and murmurs such as stenosis, diastolic, diastolic gallop etc. The classification is achieved by extracting the 2^{nd} order cumulants of the real and imaginary part of the complex envelope wave. These second-order statistical features are given to multiclass Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifier for classification. The acquired wave signals are passed through an Additive White Gaussian Noise (AWGN) channel before feature extraction. The evaluation of the system is carried on using 400 generated signals. Experimental results show that the proposed method produces an accurate classification in the range of 65% - 89% for SVM and 65% - 68% for KNN classifier.

7.1 MEL-FREQUENCY CEPSTRAL COEFFICIENTS

The spectrum of a sound signal (heart sounds and murmurs) can be considered in terms of signal correlation terms with harmonic tones at regularly spaced spectral peaks. MFCCs are a way of representing the spectral information in a sound (heart) signal. Each coefficient has a value for each frame of the sound. The sequence of steps in obtaining MFCC is:

(1) Partition the signal into frames.
(2) Get the amplitude spectrum of each frame.
(3) Compute the log of these spectrums.
(4) Convert to the Mel Scale (a perceptual scale based on human hearing).
(5) Apply the DCT.

The purpose of DCT is to reduce the data orthnormally and thereby leaving a series of uncorrelated values (the coefficients) for each frame of the heart sound signal. This is illustrated in figure 7.1

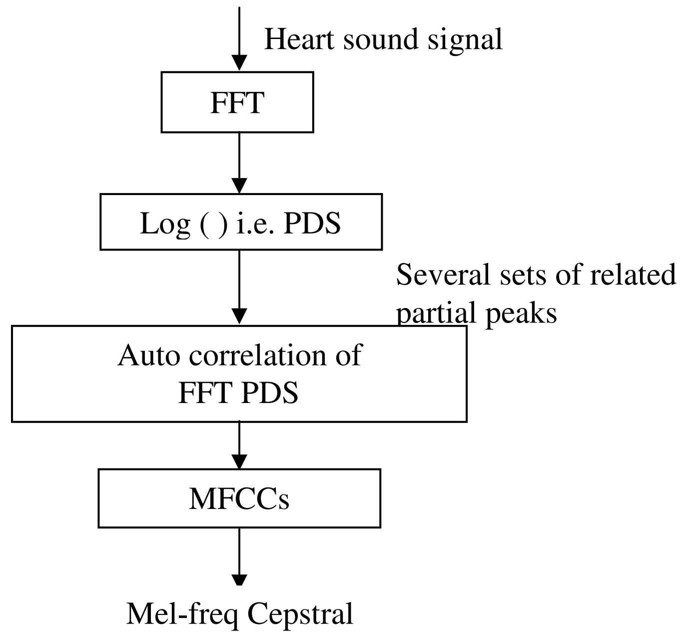


Figure 7.1 Evaluation of MFCC.

7.2 FEATURE EXTRACTION FROM HEART SOUND SIGNAL

In this research work, from the input heart sounds, two measures are used and different features are extracted. Normal and various types of abnormal heart sound signals were studied and the features (magnitude and power spectrum, FFT, zero crossing, MFCC etc.) were extracted from each case.

7.3 TRANSFORMING DATA

The data can be transformed by applying a single mathematical function to all of the observations. In the first sub-section below, they discuss the general power transformations that can be used to change the shape of the data distribution. It also covers the linear transformations of the data that leave the shape alone. These are typically changes in scale and origin and can be important in dimensionality reduction, clustering, and visualization. The order of the data is preserved by the transformation. Because of this, statistics based on order, such as medians are preserved

(i) Medians are transformed to medians.

(ii) They are continuous functions guaranteeing that points that are close together in raw form are also close together using their transformed values, relative to the scale used.

(iii) They are smooth functions that have derivatives of all orders, and they are specified by elementary functions. This is very fast to compute, so it is useful for large datasets. However, a problem with this index is that it tends to locate structure in the tails of the distribution.

7.4 DENDOGRAMS

There are several ways to visualize the output from cluster analysis and one such representation use the dendrogram to present the results. A dendrogram is a tree diagram that shows the nested structure of the partitions and how the various groups are linked at each stage. The dendrogram can be shown horizontally or vertically, however in this research the vertical version is chosen for right now, since it seems more 'tree-like.' There is a numerical value associated with each stage of the method where the branches (i.e., clusters) join, which usually represents the distance between the two clusters. The scale for this numerical value is shown on the vertical axis. The tree is made up of inverted U-shaped links, where the top of the U represents a fusion between two clusters. In most cases, the fusion levels will be monotonically increasing, yielding an easy to understand dendrogram. The problem of reversals with the centroid and median linkage methods shall exist and with reversals, these merge points can decrease, which can make the results confusing. Another problem with some of these methods is the possibility of non-uniqueness of the hierarchical clustering or dendrogram. This can happen when there are ties in the distances between clusters.

7.5 OPTIMIZATION METHODS: K–Means

The methods discussed in the previous section were all hierarchical, where the output consists of a complete set of nested partitions. Another category of clustering methods consists of techniques that optimize some criterion in order to partition the observations into a specified or predetermined number of groups. These partition or optimization methods differ in the nature of the objective function, as well as the optimization algorithm used to come up with the final clustering. One of the issues that must be addressed when employing these methods (as is also the case with the hierarchical methods) is determining the number of clusters in the data set. However, one of the major advantages of the optimization-based methods is that they require only the data as input (along with some other parameters), not the interpoint distances, as in hierarchical methods. Thus, these are usually more suitable when working with large data sets. One of the most commonly used optimization-based methods is k-means clustering. The reader is referred to Everitt, Landau and Leese [2001] for more information on the other types of partition methods. The MATLAB Statistics Toolbox has a function that implements the k-means algorithm. The goal of k-means clustering is to partition the data into k groups such that the within-group sum-of-squares is minimized. If they minimize the trace, then they can also minimizing the total within-group sum of squares about the group means. Everitt, Landau and Leese [2001] show that minimizing the trace of SW is equivalent to minimizing the sum of the squared Euclidean distances between individuals and their group mean.

Clustering methods that minimize this criterion tend to produce clusters that have a hyper-ellipsoidal shape. This criterion can be affected by the scale of the variables, so standardization should be done first. They briefly described about the two procedures for obtaining clusters via k-means. The basic algorithm for k-means clustering is a two step procedure. First, they assign each observation to its closest group, usually using the Euclidean distance between the observation and the cluster centroid. The second step of the procedure is to calculate the new centroids using the assigned observations. These steps are alternated until there are no changes in cluster membership or until the centroids do not change. This algorithm is sometimes referred to as HMEANS [Späth, 1980] or the basic ISODATA method. They note that there are many algorithms for k-means clustering described in the literature that improve the efficiency, allow clusters to be created and deleted during the process, and other improvements.

7.6 PROCEDURE FOR K - Means CLUSTERING

1. Specify the number of clusters '*k*'.
2. Determine initial cluster centroids. These can be randomly chosen or the user can specify them.
3. Calculate the distance between each observation and each cluster centroid.
4. Assign every observation to the closest cluster.
5. Calculate the centroid (i.e., the *d*-dimensional mean) of every cluster using the observations that were just grouped there.
6. Repeat steps 3 through 5 until no more changes are made.

The *k*-means algorithm could lead to empty clusters, so users should be aware of this possibility. Another issue concerns the optimality of the partitions. With *k*-means, we are searching for partitions where the within group sum-of-squares is a minimum. It can be shown that in some cases the final *k*-means cluster assignment is not optimal, in the sense that moving a single point from one cluster to another may reduce the sum of squared errors. The following procedure that we call the *enhanced k-means* helps to address the second problem.

7.7 SILHOUETTE PLOT

Kaufman and Rousseeuw [1990] present the silhouette statistic as a way of estimating the number of groups in a data set. Given observation *i*, they denote the average dissimilarity to all other points in its own cluster as a_i. For any other cluster *c*, they let $\bar{d}(i,c)$ represent the average dissimilarity of *i* to all objects in cluster *c*. Finally, they let b_i denote the minimum of these average dissimilarities $\bar{d}(i,c)$. The silhouette width for the i-th observation is

$$sw_i = \frac{(b_i - a_i)}{max(a_i, b_i)}. \qquad (7.2)$$

The average silhouette width can be obtained by averaging sw_i over all observations:

$$\overline{sw} = \frac{1}{n}\sum_{i=1}^{n} sw_i. \qquad (7.3)$$

Observations with a large silhouette width are well clustered, but those with small values tend to be ones that are scattered between clusters. The silhouette width sw_i in equation (7.2) ranges from –1 to 1. If an observation has a value close to 1, then the data point is closer to its own cluster than a neighboring one. If it has a silhouette width close to –1, then it is not very well-clustered. A silhouette width close to zero indicates that the observation could just as well belong to its current cluster or one that is near to it.

Researchers [Kaufman and Rousseeuw, (2004)] use the average silhouette width to estimate the number of clusters in the data set by using the partition with two or more clusters that yields the largest average silhouette width. They state that an average silhouette width greater than 0.5 indicates a reasonable partition of the data, and a value of less than 0.2 would indicate that the data do not exhibit cluster structure. There is also a meaningful graphical display called a silhouette plot that displays the silhouette values for each cluster, ranking them in decreasing order. This allows the analyst to rapidly visualize and assess the cluster structure.

7.8 CLASSIFIER

A study of multi-class signal classification and automatic detection through Support Vector Machines (SVM) is presented in [D. J. Liu, C. T. Lin (2001)]. A new robust classification algorithm, which applies higher-order statistics (HOS) in a generic framework for blind channel estimation and pattern recognition, is proposed in [G. Li (2003)]. Feature based method for automatic classification and recognition of hybrid waveforms is presented in [J.E. Hebden and Torry, (1996)]. The classification is conducted with artificial neural networks (ANN). The performance of energy detection based spectrum sensing for several real-world primary signals is [Grimaldi, (2003)]. A method for the automatic classification using cumulants derived using fractional lower order statistics is proposed in [Wang, (1980]. The performance of the classifier is presented in the form of probability of correct classification under noisy and fading conditions.

A novel approach based on fuzzy logic to classify signals with respect to standards on the basis of known wave parameters is presented in [T. Leung (1982)].

Ideally it would like to classify the primary user systems with respect to existing "Known standards". A novel design of Automatic parameter recognition (APR) method with reduced computational complexity and fast processing speed is needed. A discrete likelihood-ratio test (DLRT) based rapid-estimation approach to identifying the wave features blindly in real time is described in [S.L Xu, (1994)].

In this research, an approach for the heart sound signals classification based on cumulants and feature classification using Support Vector Machines (SVM) and K-Nearest Neighbor (K-NN) is presented. The methodologies and the proposed system are explained and the experimental results are explained in detail.

7.9 METHODOLOGY

The proposed system for the classification of heart murmur signals in PCG signal processing is built based on second-order statistics, along with the multiclass SVM and KNN classifier for classification.

7.9.1 SECOND – ORDER STATISTICS

7.9.1.1 AUTOCORRELATION FUNCTION AND POWER SPECTRUM

The autocorrelation function or sequence of a stationary process, $x(n)$ is defined by,

$$R_{xx}(m) = E\{ x^*(n)x(\cdot\cdot + m)\}$$

where E{·} denotes the ensemble expectation operator. The power spectrum is formally defined as the Fourier Transform (FT) of the autocorrelation sequence (the Wiener-Khintchine theorem) is given by,

$$P_{xx}(f) = \sum_{m=-\infty}^{\infty} R_{xx}(m)\exp(-j2\pi f m)$$

Where f denotes the frequency. An equivalent definition is given by

$$P_{XX}(f) := E\{X(f)X^*(f)\} \qquad 7.4$$

Where $X(f)$ is the Fourier Transform of $x(n)$

$$X(f) = \sum_{n=-\infty}^{\infty} x(n)\exp(-j2\pi fn) \qquad 7.5$$

A sufficient, but not necessary, condition for the existence of the power spectrum is that the autocorrelation be absolutely summable. The power spectrum is real valued and nonnegative, that is, $P_{XX}(f) \geq 0$; if $x(n)$ is real valued, then the power spectrum is also symmetric, that is, $P_{XX}(f) = -P_{XX}(f)$.

7.9.1.2 HIGHER ORDER CUMULANTS

The higher-order moments are natural generalizations of the autocorrelation, and cumulants are specific nonlinear combinations of these moments. The first-order cumulant of a stationary process is the mean $C_{1x} := E\{x(t)\}$. The higher-order cumulants are invariant to a shift of mean. Hence, it is convenient to define them under the assumption of zero mean. If the process has nonzero mean, then subtract the mean, apply the following definitions to the resulting process. The second-order cumulants of a zero-mean stationary process are defined as

$$C_{2x}(k) = E\{X^*(n)x(n+k)\} \qquad 7.6$$

The first-order cumulant is the mean of the process; and the second-order cumulant is the auto covariance sequence. Note that for complex processes, there are several ways of defining cumulants depending upon which terms are conjugated. The zero-lag cumulants have special names: $C_{2x}(0)$ is the variance and is usually denoted by σ_x^2.

7.10 SUPPORT VECTOR MACHINE (SVM)

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM is a non-probabilistic binary linear classifier, i.e. it predicts, for each given input, which of two possible classes the input is a member of. A classification task usually involves with training and testing data which consists of some data instances.

Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). SVM has an extra advantage of automatic model selection in the sense that both the optimal number and locations of the basis functions are automatically obtained during training. The performance of SVM largely depends on the kernel [G. Li, (2003)].

SVM is essentially a linear learning machine. For the input training sample set

$$(x_i, y_i), i = 1 \dots . n, x \in R^n, y \in \{-1, +1\}$$

the classification hyperplane equation is let to be

$$(\omega . x) + b = 0$$

thus the classification margin is 2 / |ω| .To maximize the margin, that is to minimize |ω| ,the optimal hyperplane problem is transformed to quadratic programming problem as follows,

$$\begin{cases} min\varnothing(\omega) = \frac{1}{2}(\omega, \omega) \\ s.t. y_i\big((\omega . x) + b\big) \geq 1, \;\; i = 1,2 \dots . l \end{cases} \qquad 7.7$$

After introduction of Lagrange multiplier, the dual problem is given by,

$$\begin{cases} max\, Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i . x_j) \\ s.t \;\; \sum_{i=1}^{n} y_i \alpha_i = 0, \;\; \alpha_i \geq 0, i = 1,2 \dots , n \end{cases} \qquad 7.8$$

According to Kuhn-Tucker rules, the optimal solution must satisfy

$$\alpha_i(y_i\big((w . x_i) + b\big) - 1 = 0, i = 1,2, . . n$$

That is to say if the option solution is

$$\alpha^* = \; (\alpha_1^*, \alpha_2^*, \dots . , \alpha_i^*)^T, \qquad i = 1,2, \dots n$$

Then

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i \qquad 7.9$$

$$b^* = y_i - \sum_{i=1}^{n} y_i \alpha_i^* (x_i . x_j) \ , j \in \{j | \alpha_i^* > 0\}$$

For every training sample point x_i , there is a corresponding Lagrange multiplier. And the sample points that are corresponding to $\alpha_i = 0$ do not contribute to solve the classification hyperplane while the other points that are corresponding to $\alpha_i > 0$ do, so it is called support vectors. Hence the optimal hyperplane equation is given as

$$\sum_{x_i \in SV} \alpha_i y_i (x_i . x_j) + b = 0 \qquad 7.10$$

The hard classifier is then,

$$y = sgn\left[\sum_{x_i \in SV} \alpha_i y_i (x_i . x_j) + b\right] \qquad 7.11$$

For nonlinear situation, SVM constructs an optimal separating hyperplane in the high dimensional space by introducing kernel function $K(x.y) = \emptyset(x).\emptyset(y)$, hence the nonlinear SVM is given by,

$$\begin{cases} min\emptyset(\omega) = \frac{1}{2}(\omega, \omega) \\ s.t. y_i\left((\omega.\emptyset(x_i)) + b\right) \geq 1, i = 1,2, \dots . l \end{cases}$$

And its dual problem is given by,

$$\begin{cases} maxL(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i . x_j) \\ s.t. \sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1,2, \dots, l \end{cases}$$

Thus the optimal hyperplane equation is determined by the solution to the optimal problem. A SVM classifier can predict the input data into two distinct classes. However, it can be used as multiclass classifiers by treating a K-class classification problem as K two-class problems. This is known as one vs. rest or one vs. all classification.

The SVM classifier implementation is standard implementation. In the MATLAB environment the LIBSVM software is used. LIBSVM is integrated software for support vector classification, regression and distribution estimation. It also supports multi-class classification.

7.11 K-NEAREST NEIGHBORHOOD (K - NN) CLASSIFIER

In pattern recognition, the k-nearest neighbor algorithm (K- NN) is a method for classifying objects based on closest training examples in the feature space. K - NN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until identification. In K - NN, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor. The neighbors are taken from a set of objects for which the correct identification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.

7.12 PROPOSED SYSTEM

The proposed system for the classification of heart sound murmur signals in PCG automatic heart sound signal detection mainly consists of two different phases which include the training phase and classification phase. Both these phases are explained in detail in the following sub sections. Different types of schemes are considered for the classification.

7.12.1 Training Phase

Feature extraction is a fundamental pre-processing step for pattern recognition and all machine learning problems. In the proposed method, 2^{nd} order cumulants of real and imaginary part of the complex envelope are used as features for the classification of signals. The training phase is shown in Fig 7.2.

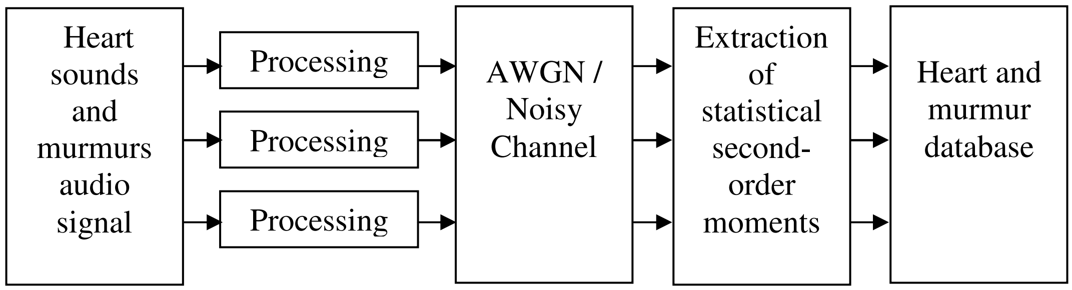


Figure 7.2 Block diagram of feature extraction phase.

The generated heart sound signals are passed through an AWGN channel with a predefined SNR level. The second-order statistical features extracted from the received signals and stored in the database for the classification purpose. The SVM and KNN classifiers are trained by using the database generated in the training phase. The algorithm is as follows.

Algorithm 7.1: Training Phase

[Input] Generated heart wave sounds

[Output] the feature vector of all signals with noise as database (DB)

1) Pass the heart sound signal through an AWGN channel with predefined SNR level
2) Calculate the 2^{nd} order cumulants by above equation 7.2
3) Step 3 is repeated for real and imaginary part of the complex envelope.
4) Insert this feature vector and the known class into the database.
5) Repeat the above steps for all heart sound murmur signals.

7.12.2 Classification Phase

In the classification phase, the unknown signal is classified. The second order statistical features are extracted from the unknown signal and this feature vector is processed with the features in the database by using the SVM and KNN classifier. The algorithm is described as follows.

Algorithm 7.2: Identification Algorithm

[Input] unknown signal and the database

[Output] the class of the signal to which this unknown signal is assigned

1) Calculate the 2nd order cumulants, shown in section 7.10
2) Step 1is repeated for real and imaginary part of the unknown signal
3) Test with the trained SVM and KNN classifier and find the class of the unknown signal.

7.13 POSITIVE PREDICTION VALUE (PPV)

In this research, the performance metric PPV is used to determine the classified accuracy of the SVM and K-NN schemes. This is defined as;

PPV = True predicted value/ (True Predicted Value (i.e. TPV) + False Predicted Value (i.e. FPV))

An ideal classifier will have a PPV value of 100% i.e. FPV = zero

7.14 EXPERIMENTAL RESULTS

7.14.1 FEATURE EXTRACTION RESULTS

In this section, the feature extracted from the heart sound corresponding to the murmur (diastolic physiologic IInd split murmur) is presented.

This includes:

(i) PCG of heart sounds and murmurs
(ii) Estimation of FFT or spectrum
(iii) Estimation of Spectrogram
(iv) MFCC histogram
(v) Dendrogram plot of the 30 murmurs set
(vi) Silhouette plots
(vii) Clusters obtained using K_p means

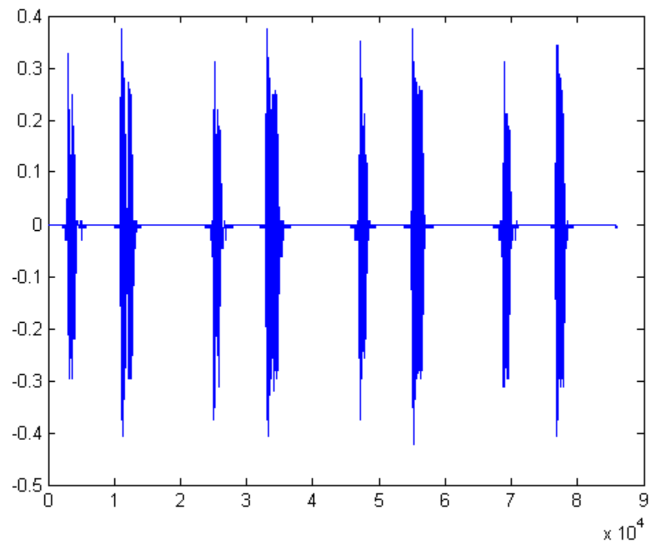


Figure 7.3 Phonocardiogram of diastolic physiologic IInd split murmur.

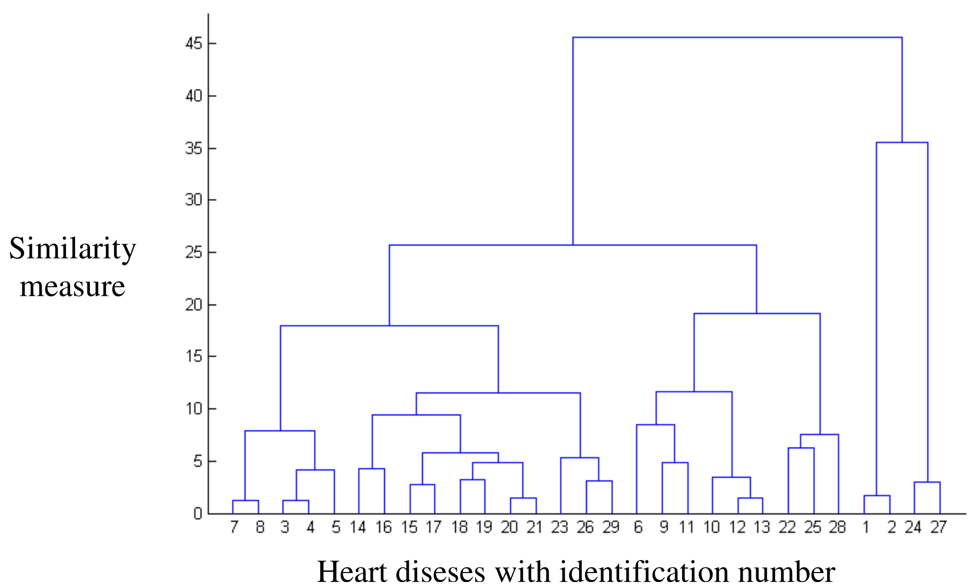


Figure 7.4 Dendrogram of the various heart diseases.

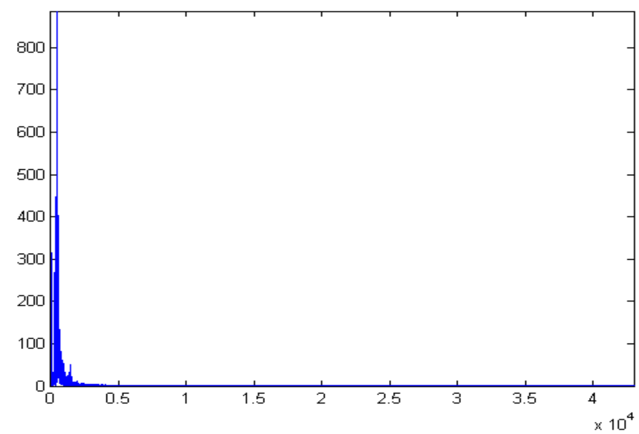


Figure 7.5 FFT spectrum of the diastolic physiologic IInd split murmur.

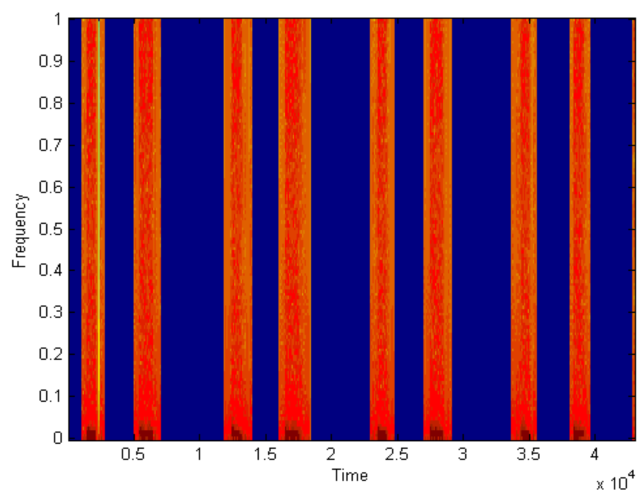


Figure 7.6 Spectrogram of diastolic physiologic IInd split murmur.

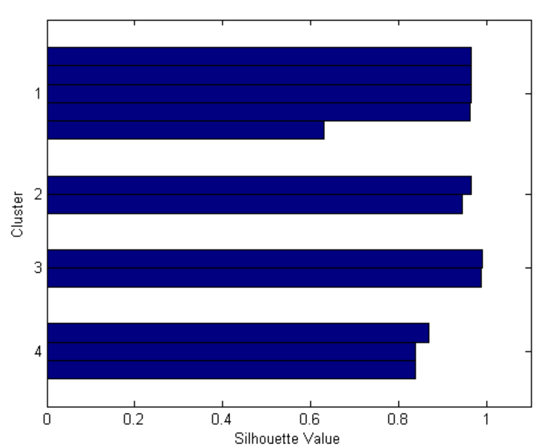


Figure 7.7 Silhouette diagram with classified clusters.

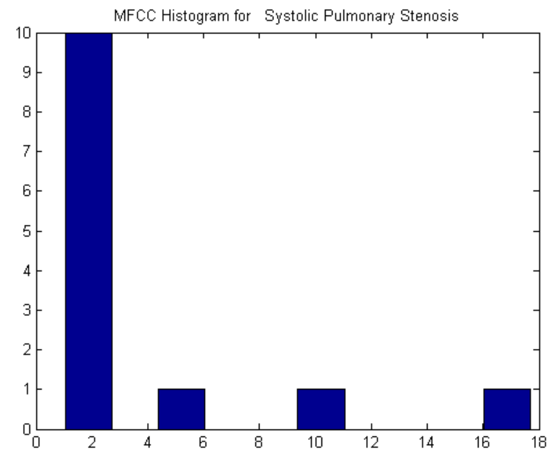


Figure 7.8 MFCC histogram plot of systolic pulmonary stenosis.

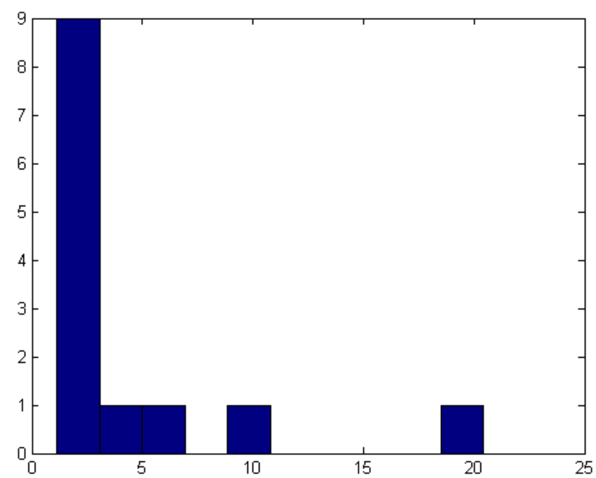


Figure 7.9 MFCC histogram plot of aortic regugration.

7.14.2 CLASSIFIER PERFORMANCE RESULTS

In this section, the performance of the proposed system based on the higher-order statistical feature discussed in this chapter is verified. A set of 400 signals with segment size of 1024 samples are generated. These 400 signals are passed through an AWGN channel of predefined SNR level. These 400 signals are separated into two set and 300 signals are randomly selected as training set and the remaining 100 signals are chosen as testing set. The SNR level used in the proposed system are 0, 1, 5 and 10 dB.

The performance metric used to evaluate the accuracy of the proposed system is the confusion matrix and the PPV. A confusion matrix represents information about actual and classified cases produced by a classification system. Performance of such system is commonly evaluated by demonstrating the correct and incorrect patterns. From Table 7.1 to Table 7.4 shows the results obtained from the SVM classifier for 1024 samples at SNR level 0, 1, 5 and 10 dB respectively and positive predictive value (PPV) also shown. From Table 7.5 to Table 7.8 shows the results obtained from the KNN classifier for 1024 samples at SNR level 0, 1, 5 and 10 dB respectively and positive predictive value (PPV) also shown. From the results it is concluded that the classification accuracy increases as SNR increases and the SVM classifier outperforms the KNN classifier. Figure 7.10 shows the overall classification accuracy of the proposed system. In table 7.1, among the 400 signals generated per each heart sound and murmur signal scheme for 0 dB, the true positive value for stenosis (ST), diastolic (DI) and ejection (EJ) are 116, 400 and 267 respectively. The overall classification rate for 1024 samples at 0, 1, 5 and 10 dB is 65.25, 73.33, 88.83 and 88.91 respectively.

Table 7.1 SVM Classification accuracy for 1024 samples at 0 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	116	0	133	46.59
Diastolic	0	400	0	100
Ejection	284	0	267	48.46
Accuracy (%)	29	100	66.75	65.25

Table 7.2 SVM Classification accuracy for 1024 samples at 1 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	213	0	133	61.56
Diastolic	0	400	0	100.00
Ejection	187	0	267	58.81
Accuracy (%)	53.25	100	66.75	73.33

Table 7.3 SVM Classification accuracy for 1024 samples at 5 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	399	0	133	75.00
Diastolic	0	400	0	100.00
Ejection	1	0	267	99.63
Accuracy (%)	99.75	100	66.75	88.83

Table 7.4 SVM Classification accuracy for 1024 samples at 10 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	400	0	133	75.05
Diastolic	0	400	0	100.00
Ejection	0	0	267	100.00
Accuracy (%)	100	100	66.75	88.92

Table 7.5 KNN Classification accuracy for 1024 samples at 0 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	180	0	196	47.87
Diastolic	0	400	0	100.00
Ejection	220	0	204	48.11
Accuracy (%)	45	100	51	65.33

Table 7.6 KNN Classification accuracy for 1024 samples at 1 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	204	0	196	51.00
Diastolic	0	400	0	100.00
Ejection	196	0	204	51.00
Accuracy (%)	51	100	51	67.33

Table 7.7 KNN Classification accuracy for 1024 samples at 5 dB SNR.

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	209	0	196	51.60
Diastolic	0	400	0	100.00
Ejection	191	0	204	51.65
Accuracy (%)	52.25	100	51	67.75

Table 7.8 KNN Classification accuracy for 1024 samples at 10 dB SNR

Heart murmur Type	Stenosis	Diastolic	Ejection	PPV
Stenosis	214	0	196	52.20
Diastolic	0	400	0	100.00
Ejection	186	0	204	52.31
Accuracy (%)	53.5	100	51	68.17

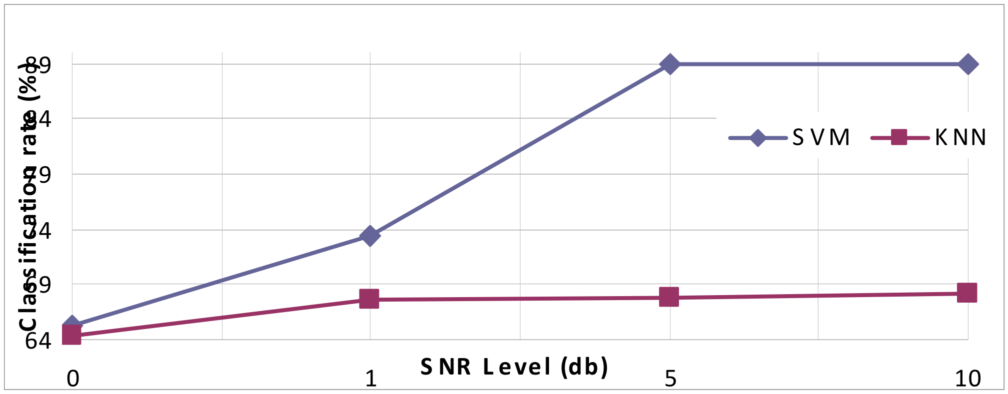


Figure 7.10 Overall classification rates (%) for 1024 samples.

7.3 CONCLUSIONS

In this chapter, feature description and an approach for multiclass signal classification based on second-order statistical features is presented. The 2ndorder cumulants of the real and imaginary part of the complex envelope are used as features for multi signal classification. The proposed system is tested on three heart murmur sound schemes. Two different classifier, SVM and KNN classifier are used to classify the digital signals. From the results, SVM classifier outperforms the KNN classifier for digital signal classification and also it is observed that the classification accuracy of the stenosis scheme for 0 dB is much lesser (i.e. poor) than all other schemes used. Confusion matrix is used to evaluate the performance of the proposed system and the experimental results prove that the proposed system provides satisfactory performance for the multi signal classification.

CHAPTER – 8

RESULTS AND DISCUSSIONS

The heart sound murmurs were recorded from different subjects along with other measured signals. The heart sound and murmurs recoding were performed using EMT25C and Meditron at various sampling frequency of 2 KHz, 2.5 KHz 6 KHz and 44.1 KHz to study the effect of sampling frequency. The subjects were both male and female with men age of 28 in set I and mean age of 65 in set II were used. The complete details are described in the Appendix – A and include both normal and abnormal heart sound recordings. From these observations, different features were extracted and these features were applied to a classifier network for further classification. The performance of the classfier was measured in terms of classification rate.

8.1 CLASSIFICATION OF HEART SOUNDS AND MURMURS

The filtered heart sound signals (wave signals) were classified using SVM technique. The classified results along with the classification/misclassification matrix for five different heart sound signals are given in Table 8.2 to Table 8.5 for different signal to noise levels. The noise levels were SNR = 0 dB, SNR =1 dB, SNR=1 dB, SNR=10 dB for 1024 samples. The classification rate is dependent on the SNR and algorithms were designed perform well with the noisy data.

The performances of the implemented algorithms were based on classification rate and also derived the classification/misclassification matrix. A confusion matrix is a specific table layout that allows visualization of the performance of a classification algorithm. For example, a sample confusion or misclassification matrix is given in Figure 8.2 (b) for a set of five heart sounds and murmurs. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another). All correct guesses are located in the diagonal of the table, so it's easy to visually inspect the table for errors, as they will be represented by any non-zero values outside the diagonal.

Table 8.1 Classification rate (a) and misclassification matrix (b) of heart sound and murmurs classified for SNR=5dB for 1024 samples.

Correctly classified diastolic rumble	100.00%
Correctly classified systolic aortic stenosis	99.25%
Correctly classified tricuspid regugration	84.00%
Correctly classified ejection murmur	94.25%
Correctly classified diastolic ventricular gallop	100.00%

(a) Classification rate of heart sound and murmurs.

400.0000	0	64.0000	0	0	86.2069
0	397.0000	0	23.0000	0	94.5238
0	0	336.0000	0	0	100.0000
0	3.0000	0	377.0000	0	99.2105
0	0	0	0	400.0000	100.0000
100.0000	99.2500	84.0000	94.2500	100.0000	95.5000

(b) Classification/misclassification matrix.

In table 8.1 (a), we can it can be observed that classification is almost 100% and is about 84% in case tricuspid regugration is mainly due to the high variability. The corresponding classification matrix is given in table 8.1 (b) and indicates high degree of correct classification.

Table 8.2 Classification rate of heart sounds and murmurs classified for SNR=0 dB 1024

Correctly classified diastolic rumble	35.25%
Correctly classified systolic aortic stenosis	72.75%
Correctly classified tricuspid regugration	98.25%
Correctly classified ejection murmur	81.75%
Correctly classified diastolic ventricular gallop	99.75%

(a) Classification rate of heart sounds and murmurs.

141.0000	0	3.0000	0	0	97.9167
0	291.0000	0	63.0000	0	82.2034
257.0000	0	393.0000	10.0000	1.0000	59.4554
2.0000	109.0000	4.0000	327.0000	0	73.9819
0	0	0	0	399.0000	100.0000
35.2500	72.7500	98.2500	81.7500	99.7500	77.5500

(b) Classification/misclassification matrix

In table 8.2 (a), it can be observed that classification rate is little bit reduced and corresponding classification matrix is given in table 8.2 (b) and indicates slight variations in classification.

Table 8.3 Classification rate of heart sound and murmurs classified for SNR=1 dB, 1024 samples.

Correctly classified diastolic rumble	80.00%
Correctly classified systolic aortic stenosis	80.00%
Correctly classified tricuspid regugration	92.50%
Correctly classified ejection murmur	82.00%
Correctly classified diastolic ventricular gallop	100.00%

(a) Classification rate of heart sound and murmurs classified.

320.0000	0	28.0000	2.0000	0	91.4286
0	320.0000	0	67.0000	0	82.6873
80.0000	0	370.0000	3.0000	0	81.6777
0	80.0000	2.0000	328.0000	0	80.0000
0	0	0	0	400.0000	100.0000
80.0000	80.0000	92.5000	82.0000	100.0000	86.9000

(b) Classification/misclassification matrix.

In table 8.3 (a), it can be observed that classification rate has improved when compared with the table 8.3 when SNR = 1 dB corresponding classification matrix is given in table 8.3 (b) and indicates better performance.

Table 8.4 Classification rate of heart sound and murmurs classified for SNR=10 dB 1024 sample.

Correctly classified diastolic rumble	99%
Correctly classified systolic aortic stenosis	100%
Correctly classified tricuspid regugration	100%
Correctly classified ejection murmur	100%
Correctly classified diastolic ventricular gallop	100%

(a) Classification rate of heart sound and murmurs

396.0000	0	0	0	100.0000
0	400.0000	0	0	100.0000
4.0000	0	400.0000	0	99.0099
0	0	0	0	100.0000
0	0	0	400.0000	100.0000
99.0000	100.0000	100.0000	100.0000	99.8000

(b) Classification / misclassification matrix

In table 8.4 (a), it can be observed that classification rate has improved significantly and classification rate is almost 100% when SNR = 10 dB and corresponding classification matrix is given in table 8.4 (b) and indicates excellent performance.

8.2 INFERENCES FROM THE RESULTS

The five different heart sound murmurs were classified from the filtered 1024 samples of the wave signal of the heart sounds and murmurs. For example, in table 8.4, corresponding to 1dB the 1st row of the classification/misclassification matrix is reproduced below:

320.0000	0	28.0000	2.0000	0	91.4286

This implies that diastolic rumble is classified correctly for 320 instants, misclassified as tricuspid regugration in 28 instants and misclassified as ejection murmur in 2 instants. The classification accuracy for diastolic rumble signal is calculated as 320/(320+28+2) = 91.4286%. Similar results are obtained for other heart sounds and murmurs. Classification plots for different heart sound murmurs are reported in the thesis. In this chapter, feature description and an approach for multiclass signal classification based on second-order statistical features is presented. The 2ndorder cumulants of the real and imaginary part of the complex envelope are used as features for multi signal classification.

The proposed system is tested on three heart murmur sound schemes. Two different classifier, SVM and KNN classifier are used to classify the digital signals. From the results, SVM classifier outperforms the KNN classifier for digital signal classification and also it is observed that the classification accuracy of the stenosis scheme for 0 dB is much lesser (i.e. poor) than all other schemes used. Confusion matrix is used to evaluate the performance of the proposed system and the experimental results prove that the proposed system provides satisfactory performance for the multi signal classification.

We also proposed psychoacoustic models based on a psychoacoustic principles and mathematical foundations and discussed the psychoacoustic features (pitch, intensity, timbre, loudness, power, intensity and other clinically important psychoacoustic features) that can be modeled, analyzed and provide effective aid of clinical decisions related to heart diseases, and in particular murmurs. These models offer a reasoning framework for the subjective reasoning of heart sounds and derived psychoacoustical models. It is also used to model the quality of heart sounds for many standardization efforts and can be used as an effective teaching aid for the cardiac auscultations. Our experimental results on psychoacoustic models are quite encouraging with average 96% classification rate when compared with spectral properties based retrieval algorithms.

A series of experiments were conducted to know the precision and retrieval efficiency. About 200 heart sounds and murmurs covering 30 types of cardiovascular pathologies were stored in the database by acquiring from the electronic stethoscope. Some of the heart sounds and murmurs were collected from the Texas Heart Research Center, USA. The heart sounds and murmur database contained the normal and benign heart sounds and innocent murmurs. Abnormal heart sounds and murmurs were compared with the normal sounds using similarity measures using MFCC. The MFCC shows remarkable distinguishing features in normal heart sound and murmurs. When comparing with murmurs, the ZCR plays critical role in separating murmurs and number of the sign change determines the murmurs vector features.

The retrieval of S3 and S4 hearts sounds was difficult and performed poorly by retrieving 21 false audio files in a collection of 10 relevant files for S3 and S4. The characterizations of S3 and S4 features vectors in terms of pitches was inadequate and recommend using other features such as magnitude spectral or spectral centroid. In our experiments, the similarity measures for detecting abnormality is calculated to be 97 % and experiments also records a detection of murmurs (early systolic and diastolic rumble) showed a better precision of 80% match to the unknown murmur. The ranked indexing with similarity measure is difficult to model using only one feature vector. For example, if we use pitch as indexing parameter and many heart sounds and murmurs file retrieved and inadequate for indexing and retrievals. The performance of the histogram search in comparison with correlation coefficient matching showed that the histogram search algorithm can achieve a high precision about 97%.

As discussed, histogram modeling using MFCC feature vectors and various techniques of similarity measures are discussed for pattern matching of PCG signals. Search accuracy of histogram matching is tested with respect to correlation algorithms and it is observed that histogram algorithm for content based heart sound retrieval is more efficient (85%) and (94%) accurate.

The distance measurement algorithms (e.g., Euclidian Distance) is computationally simple and various as function of the square root or sum squared and it may significant effect when the number of audio objects is large. However, we can use optimizes techniques such as caching and intelligent pre processing steps helps to reduce the computational complexity. The indexing time is decrease exponentially for small indexing terms and may increase exponential after a threshold value of high degree of dimensionality. ($c = 20$) or as a function of features vectors and their dimensionality.

For signals with skewed spectrum (Figure 6.2) minimum in the response time is achieved for a small number of Fourier coefficients ($f = 10, 12, 13$). The minimum is rather flat, which implies that a suboptimal choice for f will give search time that is close to the minimum and we can exploit the energy concentrating properties. The distance based histogram matching algorithm was also experimented with MFCC coefficients with dimensionality ($c = 12$) and matching algorithms gave a better results (few false alarms) and best retrieval times less than 100 ms.

8.3 FUTURE RESEARCH WORK

In the phono-cardiology research, there are pathological murmurs as well non-pathological murmurs. The non-pathological murmurs are also called benign murmurs and are commonly observed in children. The non-pathological murmurs do not have any impact on heart and are generally ignored. The current research work can be extended to study the non-pathological murmurs and devise new algorithms so that the classification accuracy is high in the presence of pathological and non-pathological origin of murmurs. The correct classification pathological murmur to a non-pathological one will assist the cardiologist for better clinical investigations.

In general, the physiological signals exhibit high degree of variability and are highly non-stationary. The retrieval and classification algorithms reported in this research work can be extended to deal with high variability in heart sounds and murmurs. The current work can be extended and derive robust algorithms which can perform well in the presence of high variability and background noise.

In this research work, it also observed that there is a close relationship between the electrocardiography (ECG) and phonocardiography (PCG) signals. In particular, the timing of cardiac events and acoustic pressure in the heart region are closely related. The ECG is electrical in nature but the PCG is of acoustical nature. In the future research work, it is recommended to explore the relationship between ECG and PCG and come up with new retrieval and classification framework integrating the latest research in ECG and PCG.

APPENDIX - A

Set	Subjects	Measured signals	f_s	Sensor	Description
I	10	ECG PCG Blood pressure Respiration	2 kHz	EMT25C	10 healthy subjects (8 male, 2 female, mean age 28 years). About 20 minutes of data (5 minutes rest, about 5 minutes hypotension, 5 minutes rest, about 2 minutes hypertension and 5 minutes rest). Reference Methods: Respiration monitored with Optovent, blood pressure monitored via an automatic oscillometric instrument or continuously via intra-arterial cannula.
II	36	ECG PCG Echocardiography	44.1 kHz	Meditron	36 patients with physiological murmurs (n=7) and various degrees of aortic stenosis (n=23) and mitral insufficiency (n=6) (19 male, 17 female, mean age 69 years all with native heart valves). Reference method: Echocardiography evaluated by expert.
III	6	ECG PCG	6 kHz	EMT25C	6 healthy subjects (6 male, mean age 28 years). Nearly 2 minutes of data recorded (30s of tidal breathing about 60s of breathing

					with continuously increasing breath volumes and 10s of breath hold). Reference method: Auscultation by expert.
IV	10	ECG PCG	2.5 kHz	EMT25C	10 healthy children (5 male, 5 female, mean age 10.5 years), 30 seconds for data recorded in a sound proof room. Reference method: PCG evaluated by expert.

REFERENCES

[001] Allan S. Jaffe, MD & Wayne L. Miller, MD (2002), “Acute myocardial infarction”, Divisions of Cardiology and Laboratory Medicine, Mayo Clinic, Rochester, Minnesota, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 102-148

[002] A. Khokhar, G. Li (2000), “Content-based indexing and retrieval of audio data using wavelet” -ICME 2000, New York, pp. 234-237.

[003] A. Leatham, (1975) “Auscultation of the heart and phonocardiography”, Churchill Livingstone, London, Edinburgh.

[004] A. Leatham, (1957) “Auscultation and phonocardiography: a personal view of the past 40 years”, British Heart Journal, May, 55 (5): pp. 316-7.

[005] A.A Luisada, (1965) “Auscultation and phonocardiography”, Mosby, St. Louis, USA.

[006] A Ghias, J. Logan, and D. Chamberlain. B. C. Smith (1995), “Query by humming.” In Proc.- ACM Multimedia 95, pp. 231-236.

[007] A. LP Chen, M. Chang and J. Chen. (1999),"Query by music segments: an efficient approach for song retrieval". - In the Proc. of ICME 2000, Berlin, pp. 453-459.

[008] A Yoshitaka and T. Ichikawa.(1999), “A survey on content-based retrieval for multimedia databases.” -IEEE Trans. Knowledge and Data Engineering, 11(1):pp. 81-93.

[009] A. Uitdenbogerd and J. Zobel., (1999), “Melodic matching techniques for large music database.” -In *Proc.* ACM Multimedia 99, pp. 57-66.

[010] A. K. Jain and R. C. Dubes (1988), “Algorithms for clustering data” Prentice-Hall. London.

[011] A. Klapuri, (1999), “Sound onset detection by applying psychoacoustic knowledge”,- in Proc. IEEE Conf., Acoustics, Speech and Signal Processing, Paris, pp. 253-260..

[012] A. P. Klapuri, A. J. Eronen, and J. T. Astola, (2006),”Analysis of the meter of acoustic musical signals” - IEEE Trans. Speech Audio Proc., vol. 14, no. 1, pp. 1024-1030.

[013] A. Wang, (2006),”The Shazam music recognition service”, Communication of the ACM, vol. 49, no. 8, pp. 44–48.

[014] A. Freed, (2006),”Music metadata quality: A multiyear case study using the music of Skip James”,- in Proc. AES 121st Conv., San Francisco, CA, pp. 765-770..

[015] A. Edgecliffe-Johnson. (2006), ‘‘Rivals aim at iTunes’ grip on music downloads,’’ - Financial Times. [Online]. Available: www.ft.com

[016] A. N. Pelech (2004), “The physiology of cardiac auscultation”, Pediatric Clinical of North America, Vol (51), pp. 1515-1535.

[017] Ara G. Tilkian and Mary Boudreau Conover (2001), “Understanding heart sounds and murmurs: with an introduction to lung sounds”. Saunders, Philadelphia, 4. Edition

[018] B.G. Wells, Rapport M.B., Sprague H.B. (1949), "The sounds and murmurs in contraction of the aorta: a study by auscultation and phonocardiography", American Heart Journal, Jul. 38(1), pp. 69-79.

[019] B.Erickson, Heart Sounds and Murmurs,"A practical guide". Mosby-year Book Inc., St. Louis, 1997, USA.

[020] Blase A. Carabello, MD & Michael H. Crawford, MD (2002), "Aortic stenosis", Mayo Clinic, Scottsdale, Arizona, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 185-205.

[021] Brian D. Hoit, MD & Michael D. Faulx, MD (2002), "Tricuspid and pulmonic valve disease", Professor of Medicine, Case Western Reserve University and University Hospitals of Cleveland, Ohio, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 254-273.

[022] B. Shipton, W. Wahba (2001), "Valvular heart disease: review and update", American Family Physician, 63(11), pp. 2201-2208.

[023] B.C.J More, B.R. Glasberg and T. Baer (1997), "A model for the prediction of thresholds, loudness and partial loudness", Journal of the Audio Engineering Society, vol. 45, no. 4, pp. 224-240.

[024] B.R Glasberg and B.C Moore, "A model of loudness applicable to time varying sounds", Journal of the Audio Engineering Society, Vol. 50, No. 5, pp. 331-342, 2002

[025] Blum, T., Keislar, D., Wheaton, J., and Wold, E. (1997), "Audio databases with content-based retrieval," - Intelligent Multimedia Information Retrieval, AAAI Press, Menlo Park, California, pp.113-135,

[026] B. Logan and A. Salomon, (2001),"A music similarity function based on signal analysis",- Proc. IEEE Int. Conf. Multimedia and Expo, Tokyo, Japan, pp. 876-882,.

[027] Baeza-Yates, and Gonet (1992), "Fast text searching allowing errors" Communication of the ACM, Vol. 35(10), pp. 74-82.

[028] Bainbridge, D. (1997) Extensible optical music recognition. PhD thesis, Department of Computer Science, University of Canterbury, New Zealand., 1997.

[029] B. Bell and B. Vecchione (1993), "Computational musicology", Computers Humanities, vol. 27 (1).

[030] Blase A. Carabello, MD & Michael H. Crawford, MD (2002), "Aortic stenosis", Mayo Medical School, Rochester, Minnesota, Consultant in Cardiovascular Diseases, Mayo Clinic, Scottsdale, Arizona, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 185-205.

[031] B. Whitman and R. Rifkin (2002), "Musical query-by-description as a multiclass learning problem",- in Proc. IEEE Multimedia Sig. Proc. Conf., St. Thomas, USA, pp. 763-770.

[032] C. Guyton Arthur and E. Hall John (2005), "Textbook of medical physiology" Saunders W.B.

[033] Craig Timm, MD (2002) , “Cardiogenic shock”, University of New Mexico Health Sciences Center, Albuquerque, Current Diagnosis and Treatment in Cardiology 2nd Ed, pp.149-163.

[034] Chad Stoltz, MD & Robert J. Bryg, MD (2002), “Mitral stenosis”, Department of Cardiovascular Medicine, University of Wisconsin School of Medicine, Madison, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 224-238,

[035] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, (2003) “Complex domain onset detection for musical signals”, - in Proc. Int. Conf. Digital Audio Effects, London, U.K, pp. 746-752.

[036] C. Yang (2001), “Music database retrieval based on spectral similarity”,-Stanford University Database Group Technical Report pp. 2001-2014.

[037] C. Shub (2003), “Echocardiography or auscultation? How to evaluate systolic murmurs. Can Fam Physician”, Vol (49), pp. 163–167.

[038] D.H.Bekkering, Weber J.(1994) , Standardization of phonocardiography: effort in the Netherlands, American Heart Journal, Aug, 54 (2):pp. 316-322.

[039] D.D. Greenwood (1961), “Critical bandwidth and the frequency coordinates of the basilar membrane,”- Journal of Acoustic Society of America pp. 1344-1356.

[040] D. P. W. Ellis and G. E. Poliner (2007), “Identifying cover songs_ with chroma features and dynamic programming beat tracking” ,- in Proc. Int. Conf. Acoustic, Speech and Signal Processing, Honolulu, pp. 736-742.

[041] D. Byrd, “A similarity scale for content-based music IR”, [Online]. Available: www. informatics.indiana.edu/donbyrd/ MusicSimilarityScale.html. (last accessed Dec. 2011)

[042] D. Tjondronegoro, Y. Chen, and B. Pham (2004), “Applications: The power of play-break for automatic detection and browsing of self-consumable sport video highlights”,- In Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pp. 267–274.

[043] D. Reynolds and R. Rose (1995), “Robust text-independent speaker identification using Gaussian mixture speaker models”,- In IEEE Transactions in Speech and Audio Processing, Vol 3: pp. 72–83.

[044] D. Mitrovic and M. Zeppelzauer. (2006), “Discrimination and retrieval of animal sounds”- In Proceedings of the IEEE Conference on Multimedia Modelling, Berlin, pp. 971-981.

[045] D. Comer (1979), "The Ubiquitous B-Tree". ACM Computing Surveys, Vol.11 (2): pp. 121-128.

[046] D. A. Grossman and O. Frieder (1998), "Information retrieval: algorithms and heuristics".

[047] De Poli, G., Prandoni, P (1997), “Sonological models for timbre characterization”. J., New Music Research, Vol.26, pp.170-197.

[048] D. J. Liu, C. T. Lin (2001), “Fundamental frequency estimation based on the joint time-frequency analysis of harmonic spectral structure”- IEEE Transactions on Signals and Systems, pp. 609-621.

[049] D.D. Greenwood (1961), “Critical bandwidth and the frequency coordinates of the basilar Membrane,” Journal Acoustics Society of America, pp. 1344 -1356.

[050] Daniel and R. Weber (1997), “Psychoacoustical roughness: implementation of an optimized model”, Acustica, Vol. 83, pp. 113-123.

[051] E.Mannheimer (1956), “Standardization of phonocardiography”, Proc.of Second European Congress of Cardiology, Amsterdam, pp. 865-872.

[052] E. Zwicker and H. Fastl (1990), “Psychoacoustics facts and models”, Springer-Verlag, Berlin.

[053] E. Zwicker and U. Zwicker (1991), “Audio engineering and psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System,” J. Audio Eng. Soc., pp. 115-126.

[054] E. Schreirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’97), pp. 1331-1334.

[055] E. Clarke and N. Cook (2004), “Empirical musicology: aims, methods, prospects”-. New York: Oxford University. Press.

[056] E. Selfridge-Field, (1998), “Conceptual and representational issues in melodic comparison “, - in Melodic Similarity Concepts, Procedures, and Applications, W. B. Hewlett and E. Selfridge Field, Eds. Cambridge, MA: MIT Press.

[057] E. Gomez and E. Chew, Eds., (2006), “Tonal description of polyphonic audio for music content processing ”, - INFORMS J. Computing, Special Cluster Computation in Music, vol. 18, no. 3, pp. 826-836.

[058] E. Scheirer and M. Slaney (1997), “Construction and evaluation of a robust multifeature speech/music discriminator”,- in Proc. Int. Conf. Acoustics, Speech, Signal Processing, Munich, Germany, pp. 972-982.

[059] E. Pampalk, (2006), “Computational models of music similarity and their application in music information retrieval”, - Ph.D. dissertation, Vienna Univ. Technology, Vienna, Austria.

[060] E. Pampalk. (2004), “A matlab toolbox to compute similarity from audio”, - In Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR’04), 2004, pp. 973-983.

[061] E. Wold (1996), “Content-based classification, search and retrieval of audio data,” IEEE Multimedia, vol. 32, No. 4, pp. 426-434.

[062] Eronen, A., Klapuri, (2000), “Musical instrument recognition using cepstral coefficients and temporal features”, In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 753-756.

[063] E. D. Scheirer, (1998), “Tempo and beat analysis of acoustic musical signals”,- J. Acoust. Soc. Amer., vol. 103(1), pp. 588–601.

[064] Eronen, (2003), “Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMS”,- In ISSPA -03, Frankfurt, pp. 987-994.

[065] Fujishima, (1999), "Real-time chord recognition of musical sound: A system using commonlisp music",- in Proc. Int. Computer Music Conf., Beijing, China, pp. 765-770.

[066] G.V Bismark (1974), "Sharpness as an attribute of the timbre of steady sounds", Acustica, vol. 30, pp. 159-172.

[067] G. Tzanetakis and P. Cook (2000), "Audio information retrieval (AIR) Tools", in International Symposium on Music Information Retrieval, London, pp. 643-649.

[068] G. Li and Z. Li. (2003).Content-based classification and retrieval by support vector machines. In IEEE Transactions on Neural Networks, Vol.14: pp.209-215.

[069] G. Salton and M. McGill (1983), "Introduction to modern information retrieval", - New York McGraw-Hill, USA.

[070] H. Kim, N. Moreau, and T. Sikora (2004), "Audio classification based on mpeg- 7 spectral basis representations ", In IEEE Transactions on Circuits and Systems for Video Technology, pp. 716–725.

[071] H.Liang, S.Lukkarinen and I Hartimo (1999), "Heart sound segmentation algorithm based on heart sound envelogram, computational cardiology", Vol.24, pp.105-108.

[072] H. Ljunggren (1949), "Investigation of an apparatus for calibrated phonocardiography according to the Mannheimer-Stordal system", Acta Med Scand, Jun 30, Vol.133 (6): pp.388-93.

[073] H. Fletcher (1940), "Auditory patterns," Rev. Mod. Phys., pp. 47-65.

[074] Herrera, Pamatriain (2000), "Towards instrument segmentation for music content description", A Critical View of Instrument Classification Techniques. In ISMIR -2000, pp.456-462.

[075] H. Vermarien (2006), "Phonocardiography". In J. G. Webster, editor, Encyclopedia of medical devices and instrumentation. Wiley-Interscience, Wiley and Sons, Hoboken, N.J,", pp. 278-290

[076] I. Daubechies (1990), "The wavelet transform, time-frequency localization and signal analysis", IEEE Transactions on Information Theory, Vol.36 (5). pp. 961-1005.

[077] J. T. Foote (1997), "Content-based retrieval of music and audio.",- In Proc. SPIE, Vol. 3 (229), pp. 138-147.

[078] J. Foote (1999), "An overview of audio information retrieval." In ACM Multimedia Systems, Vol. 13, No. 6, pp 2-10.

[079] J. Foote. (1999), "Visualizing music and audio using self-similarity." In Proc. ACM Multimedia *99*, pp 77-80.

[080] J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner (1976), "Real-time digital hardware pitch detector.", IEEE Trans. On Acoustics, Speech, and Signal Processing, ASSP-Vol. 24(1): pp. 2-8.

[081] J. Johnston (1988), "Estimation of perceptual entropy using noise masking criteria," in Proc. ICASSP- 88, pp. 2524-2527.

[082] J. Zwislocki (1965), “Analysis of some auditory characteristics,” in Handbook of Mathematical Psychology, R. Luce, et al., Eds., New York: John Wiley and Sons, Inc.,

[083] J. Serra and E. Gomez, (2007), “A cover song identification system based on sequences of tonal descriptors”, - in Proc. Int. Conf. Music Information Retrieval, Vienna, Austria, pp. 762-769.

[084] J.Elizabeth Corwin (2006), “Handbook of pathophysiology: foundations of health and disease”, 3rd Ed., London.

[085] J.E.Hebden, J.N.Torry (1996), “Neural network and conventional classifiers to distinguish between first and second heart sounds. Artificial Intelligence Methods for biomedical Data Processing”, IEEE colloquium, April, London, Vol.3, pp.1-6.

[086] J. S. Downie (2006), “The music information retrieval evaluation exchange (MIREX)”, - D-Lib Mag., vol. 12 .pp 12-20.

[087] J. S. Downie, K. West, A. Ehmann, and E. Vincent (2005), “The 2005 music information retrieval evaluation exchange”, - in Proc. Int. Conf. Music Information Retrieval, London, U.K, pp. 753-760.

[088] J. C. Brown and M. S. Puckette, (1992), “An efficient algorithm for the calculation of a constant Q transform”, - J. Acoust. Soc. Amer., vol. 92, no. 5, pp. 2698–2701.

[089] J. P. Christopher (2005), “The Sense of hearing”, 2nd ed., Rutledge, New York,.

[090] J. Katz, R.F. Burkard and L. Medwetsky (2002), Handbook of clinical audiology, Lippincott Williams and Wilkins, , New York.

[091] J. Atallah, F. Chyzak, P. Dumas (2000), “A randomized algorithm for approximate string matching” Algorithmica. pp. 468-486.

[092] J.Ristama. (1967), “Frequencies dominant in the perception of the pitch of complex sounds” J. Acoustic Soc. America, Vol.24 (1): pp.191-198.

[093] Kathleen M. Allen, MD (2002), “Coronary revascularization”, The University of New Mexico School of Medicine, Albuquerque, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 164-184.

[094] Keislar, D., Blum, T., Wheaton, J., and Wold, E (1995)., "Audio databases with content-based retrieval," Proc. of the International Computer Music Conference 1995, pp. 199-202.

[097] Khokhar (2000), G. Li “Content-based indexing and retrieval of audio data using wavelet” ICME, New York, pp. 276-282.

[098] Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima(2000), “A practical query-by-humming system for a large music database.”, In Proc. of the 8th ACM International Conference on Multimedia, pp. 273-280.

[099] Kosugi, Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima (1999), “Let's search for songs by humming!” In Proc. ACM Multimedia 99 (Part 2), pp 194-200.

[100] K. Kushima, N. Kosugi, Y. Nishihara, S. Kon'ya, M. Yamamuro,. (1999), “Music retrieval by humming.” In Proc. Of PACRIM'99, pp. 404-407.

[101] Kageyama, Y. Takashima (1994), "A melody retrieval method with hummed melody", Journal of The Institute of Electronics Information and Communication Engineers, pp.1543-1551.

[102] Kemp (1995), "Multimedia and spatial information systems," IEEE Multimedia, Vol. 2(4), pp. 289-296.

[103] K. Dumont (2004), "Experimental and numerical modeling of heart valve dynamics", PhD thesis, Gent, Belgium

[104] Logan. B (2000), "Mel frequency cepstral coefficients for music modeling". In proceedings ISMIR – 2000, pp. 387-296.

[105] L. Lu, H. You, H. J. Zhang (2001), "A new approach to query by humming in music retrieval" in ICME2001, Tokyo, pp. 987-997.

[106] L. Kjell and L. Pauli (1998), "Musical information retrieval using musical parameters" International Computer Music Conference, Ann Arbour, pp. 876-883.

[107] L. S¨ornmo and P. Laguna (2005), "Bioelectrical signal processing in cardiac and neurological applications", Elsevier, New York

[108] Lee, Y.M.; Moghavvemi, M, "Remote heart rate monitoring system based on phonocardiography", Research and Development, 2002. SCOReD 2002. pp. 823-832.

.

[109] Michael H. Crawford, MD (2002) , "Approach to cardiac disease diagnosis", Mayo Medical School, Rochester, Minnesota, Mayo Clinic, Scottsdale, Arizona, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 14-36.

[110] Michael H. Crawford, MD (2002) , "Chronic ischemic heart disease", Mayo Medical School, Rochester, Minnesota, Mayo Clinic, Scottsdale, Arizona, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 61-81.

[111] Michael H. Crawford, MD (2002), "Mitral regurgitation", , Mayo Medical School, Rochester, Minnesota, Mayo Clinic, Scottsdale, Arizona (1998), Current Diagnosis & Treatment in Cardiology 2nd Ed, pp.239-253.

[112] Muscle Fish LLC. *http://www.musclefish.com*

[123] McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L., and Cunningham, S.J. (1996), "Towards the digital music library: tune retrieval from acoustic input." In Proceedings of ACM Digital Libraries, 1996, pp. 1118-1124

[124] M. Charikar C., Chekuri, T. Feder and R. Motwani (1997) , "Incremental clustering and dynamic information retrieval" Proc. of ACM Symposium on Theory of Computing, pp. 687-694.

[125] M. Filickner, H. Sawhney, W. Niblack, J. Ashley, W. Huang, B. Dom, M. Gorkani, J. Hafine, D. Lee, D. Petkovic, D. Steel, and P. Yanker. (1995), "Query by image and video content: The QBIC system". IEEE Computer Society Press, 28:pp.23-32.

[126] M. Grimaldi, Cunningham P., and A. Kokaram (2003). "A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques"- In Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, pp. 102–108.

[127] M. Goto (2001), “An audio-based real-time beat tracking system for music with or without drum-sounds”, Journal of New Music Research, vol. 30, no. 2, pp. 159-171.

[128] M.S. Obiadt, Phonocardiogram signal analysis: techniques and performance comparison, “journal of medical engineering and technology”, Vol.17 (6).pp.221-232.

[129] M. A. Bartsch and G. H. Wakefield, (2001), “To catch a chorus: Using chroma-based representations for audio thumb nailing”,- in Proc. IEEE Workshop Applications of Signal Processing to Audio and Acoustics, pp. 15–18.

[130] Martin, K. (1998), “Toward automatic sound source recognition: identifying musical instruments”, Proceedings NATO Computational Hearing Advanced Study Institute, New York, pp. 743-754

[131] M. Ramı´rez, Celma and P. Herrera, (2005), “Surfing the music: A music recommendation system based on RSS feeds and user preferences”,- in Proc. 6th Int. Conf. Music Information Retrieval, London, U.K, pp. 864-870.

[132] N. Hu and R. Dannenberg, (2002), “A comparison of melodic database retrieval techniques using sung queries”, - in Proc. ACM Joint Conf. Digital Libraries, pp. 301-307.

[133] O’Shaughnessy, D (1987), “Speech Communication Human and Machine”, Addison-Wesley Series in Electrical Engineering, Washington, USA.

[134] Peter C. Chien, MD & William H. Frishman, MD (2002), “Lipid disorders”, Westchester Medical Center, Valhalla, New York, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp 37-60.

[135] Prediman K. Shah, MD & Kuang-Yuh Chyu, MD (2002), “Unstable angina”, Shapell Atherosclerosis Research Center, Cedars-Sinai Medical Center; Professor of Medicine, University of California, Los Angeles, Current Diagnosis & Treatment in Cardiology 2nd Ed, pp. 82-101.

[136] P.Stein, (1981), A physical and physiological basis for the interpretation of cardiac auscultation: evaluations based primarily on the second sound and ejection murmurs, Futura Publishing Company.

[137] P. Findlen (2011) “A history of the heart. Internet online”.

http://www.stanford.edu/class/history13/earlysciencelab/body/heartpages/.(last accessed Oct 2011).

[138] Pena C (1979) “Comparison of abdominal ECG and phonocardiography for instantaneous fetal heart rate detection IEEE Engineering in Medicine and Biology Society”,. 17th Annual Conference, pp. 786-784.

[139] P.D. Stein, H.N. Sabbah, J.B. D.J. Magilligan, Jr., and S.Goldstein (1981), “Frequency of the first heart sound in the assessment of stiffening of mitral bioprosthetic valves, circulation”, Vol.63, pp.200-203.

[140] P.Wang, C.S. Lim, S.Chauhan, Foo, J.Y.A., and Anantharaman, V., Phonocardiographic Signal Analysis Using a Modified Hidden Markov Model. Ann. Biomedical Eng., Vol.35 (3).pp. 367- 364.

[141] Rabiner and B. Juang (1993). Fundamentals of speech recognition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA,

[142] Rabiner and R. Schafer (1978). Digital processing of speech signals. Prentice Hall Inc., Englewood Cliffs, New Jersey,

[143] Remco C. Veltkamp, Mirela Tanase, and Danielle Sent (1999), "Features in content-based image retrieval systems: a survey", in State-of-the-Art in Content-Based Image and Video Retrieval. pp. 970-998.

[144] R. Subramanya and A. Youseef (1998), "Wavelet-based indexing of audio data in audio/ multimedia databases" 4th IEEE International Workshop on Multimedia DBMS, Dayton, Ohio, pp. 987-996.

[145] R.J. Rangayyan, R.J.Lehner (1988), phonocardiogram signal analysis: a review, Durand L.G., Pibarot , CRC Critical Reviews in Biomedical Engineering Vol.15, pp. 211-236.

[146] R. B. Northrop (2002), "Noninvasive instrumentation and measurement in medical diagnosis", CRC Chapman & Hall, England.

[147] R. A. O'Rourke, V. Fuster, R. W. Alexander, R. Roberts, S. B. King, I. Nash, and E. N. Prystowsky. Hurst's (2005), "The heart manual of cardiology. McGraw-Hill Professional", New York, 11th Edition.

[148] Rabiner, L. and Juang, B (1993), "Fundamentals of speech recognition, Prentice-Hall".

[149] R. M. Berne and M. N. Levy (1997), "Cardiovascular physiology. Mosby, St. Louise, Missouri", 7. Edition.

[150] S.S Stevens, J. Volkmann and E. Newman (1937), "A scale for the measurement of psychological magnitude of pitch", Journal of Acoustical Society of America, Vol. 8, pp. 185-190.

[151] Schroeder, et al. (1979), "Optimizing digital speech coders by exploiting masking properties of the human ear," J. Acoust. Soc. Am., pp. 1647-1652.

[152] Sebe, N., Lew, M.S., Zhou, X., And Huang, T.S. (2003), "The state of the art in image and video retrieval. In Proceedings of the 2nd International Conference on Image and Video Retrieval, Urbana, July 2003, E.M. BAKKER, T.S. HUANG, M.S. LEW, N. SEBE, AND X. ZHOU, Eds. Springer-Verlag, London, pp. 987-997.

[153] S.L.Xu, M.Hu, D.C.Zhang (1994), "The time-frequency representation of PCG", proceedings of the IEEE International Conference, Los Angeles, USA, pp. 679-681.

[154] S. Wu and U. Manber (1996), "Fast text searching allowing errors." Communications of the ACM, 35(10), pp.83-91

[155] S. Douglas. (2009). U.K. music downloads reach record high. [Online]. Available: www.investmentmarkets.co.uk (last accessed Oct, 2009)

[156] S. Dixon (2001), "Automatic extraction of tempo and beat from expressive performances",- J. New Music Res., Vol. 30(1), pp. 39–58.

[157] Scheirer, E. and Slaney, M., (1997), "Construction and evaluation of a robust multifeature speech/music discriminator", -Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 654-668.

[158] Sethares and T. W. Staley (2001), “Meter and periodicity in musical performance”,-J. New Music Res., Vol. 30(2), pp. 149-158.

[159] Sethares and R. Arora (2007), “Equilibrium of adaptive wavetable oscillators with applications to beat tracking”, - in Proc. Int. Cont. Acoustics, Speech Signal Processing, Honolulu, pp.987-996.

[160] Sun. (2002), “Pitch determination and voice quality analysis using subharmonic to-harmonic ratio”- In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’02), pp. 876-884.

[161] T.A.D Michael (1997), “Auscultation of heart “, .McGraw-Hill, New York, USA.

[162] TC, ALP Chen, and CC Liu (1996), "Music databases: indexing techniques and implementation," in Proc. of IEEE Intl. Workshop Multimedia Data Base Management System, pp. 876-886.

[163] T. Tremain (1982), “The government standard linear predictive coding algorithm: Lpc-10”- In Speech Technology Magazine, pp. 40-49.

[164] T.Leung, P.White. W.Collis, E.Brown and A Salmon (1982), Classification of heart sounds using time-frequency method and artificial neural networks, in proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol.2.pp. 988-991.

[165] N. Roussopoulos (1995), Steve Kelley, and F. Vincent. “Nearest neighbor’s queries”. In proc. of ACM-SIGMOD, San Jose, May, 1995, pp. 71-79.

[166] T.H.Joo (1982), “pole-zero modeling and classification of PCG”, M.Sc. thesis, Dept Electrical Engineering and Computational Science, Massachusetts institute of Technology, Cambridge, MA

[167] V. Raghavan and Gudivada (1995), “Special issue on content-based image Retrieval IEEE Computers, Vol. 28, (9), pp. 987-998.

[168] Volker G and Günter. O (1998), Multidimensional access methods, ACM Computing Surveys, 30(2): pp. 170-231

[169] N. Beckmann, H. Kriegel, R. Schneider and B.Seeger (1990). The R* tree: An efficient and robust access method for points and rectangles. ACM SIGMOD, 1990, pp. 322-331.

[170] Wold, E., Blum, T., Keislar, D., and Wheaton, J. (1996), "Content-based classification, search and retrieval of audio," IEEE Multimedia, 3(3), pp.27-36.

[171] X. Wu and M. Li (2007), “A QBSH system based on a three-level melody representation”, - in Proc. Int. Conf. Music Information Retrieval, Vienna, Austria, pp. 982-992.

[172] Y. Wang, Liu, J. Huang, and T. Chuan (1997), “Audio feature extraction and analysis for scene classification” - In IEEE Workshop on Multimedia Signal Processing, pp.343-348.

[173] Y. Rolland, G. Raskinis, and J. G. Ganascia. (1999), “Musical content-based retrieval: an overview of the melodiscov approach and system.” In Proc. ACM Multimedia 99, pp. 81-84.

[174] Y. Manolopoulos, C Faloutsos, N. Ranganathan (1994). Fast subsequence matching in time-series databases. In Proc. ACM SIGMOD, pp. 419-429.

[175] Y. Guttmann, R-trees (1984): A dynamic index structure for spatial searching. In proc. ACM SIGMOD, Boston, pp. 47-57.

[176] Y-T. Zhang, G. Chan, X-Y. Zhang, and L. Yip. Heart sounds and stethoscopes (2006), “In M. Akay, editor, Wiley Encyclopedia of Biomedical Engineering”, Pg. 1824–1834. Wiley-Interscience, Wiley and Sons, Hoboken, N.J.

[177] Yihua Xiao (1998) , “The phonocardiogram exercise test engineering ”, IEEE Magazine, Vol. 42, No. 122, pp. 596-564.

[178] Z. Kemp (1995) , “Multimedia and spatial information systems,” IEEE Multimedia, Vol. 2, No. 4, 1995, pp. 876-888.

[188] Z. More, B.R. Glasberg and T. Baer,(1976), “A model for the prediction of thresholds, loudness and partial loudness”, Journal of the Audio Engineering Society, vol. 45, no. 4, pp. 224-240.

LIST OF PUBLICATIONS

International Journal Papers

1. Kiran Kumari Patil, Dr. B.S. Nagbhushan and Dr. Vijaya Kumar B.P. "*An efficient modeling technique for heart sounds and murmurs*", published in International Journal of Scientific & Engineering Research, Volume 2, Issue 2, and February-2011 ISSN 2229-5518 IJSER.

2. Kiran Kumari Patil, Dr. B.S. Nagbhushan and Dr. Vijaya Kumar B.P. " *An efficient retrieval technique for heart sounds using psychoacoustic similarity*", Published by International Journal of Engineering Science and Technology, IJEST Vol. 2 (12), Dec 2010 Issue, pp.7324-7328

3. Kiran Kumari Patil, Dr.B.S. Nagbhushan Dr.Vijaya Kumar B.P. "*PCG-XML: A markup language and tools for phonocardiography*" International Journal of Computational Intelligence and Healthcare Informatics (IJCIHI), ISSN: 0973-7413", Vol. 2, No. 1**,** pp 119-123, Jan-June, 2009, (http://www.researchbible.org/?date=20101216)

International Conference

1. ***Kiran Kumari Patil, Dr. B.S. Nagbhushan and Dr. Vijaya Kumar B.P."*A psychoacoustic framework and user interface tool for retrieval of heart sounds and murmurs*", International Conference on Medical and Biological Engineering (ICMBE 2012), Dec-10-11, 2011, MIT, Manipal, Karnataka.***

2. ***Kiran Kumari Patil, Dr. B.S. Nagbhushan and Dr. Vijaya Kumar B.P.,*** **"Psychoacoustic models for heart sounds",** ***presented and published in Proceedings of WIMoNE 2011 / COSIT 2011, Springer publication held on JAN 2-4 2011, pp 556-563, Bangalore, India.***

3. Kiran Kumari Patil, Dr.B.S. Nagbhushan, Dr.Vijaya Kumar B.P, "*Identification of heart sounds using Mel-frequency ceptral coefficients*", presented and published in proceedings of International Conference on Systemic, Cybernetics and Informatics, ICSCI 2010, 5-7 Jan 2011, pp. 475-478**,** Hyderabad

4. Kiran Kumari Patil, Dr.B.S. Nagbhushan, Dr. Vijaya Kumar B.P. "*Content based heart sound retrieval*", presented and published in proceedings of International

Conference on Systemic, Cybernetics and Informatics ICSCI 2010, 5-7 Jan 2011, pp. 489-491,Hyderabad.

4. Kiran Kumari Patil, Dr.B.S. Nagbhushan, Dr.Vijaya Kumar B.P, "*A markup language - mPCG-xml for mobile health care systems*", Proceedings of 7th International Conference on Information, Communication, and Signal Processing, Macau, December, 2009, ISBN: 978-1-4244-46568.

5. Kiran Kumari Patil, Dr.B.S. Nagbhushan Dr.Vijaya Kumar B.P., and, "*PCG-XML: A markup language and tools for phonocardiography*", International Conference on VLSI and Communication ICVCOM-2009, 1-19 April 2009, Kottayam, Kerala, India.

National conference

1. Kiran Kumari Patil, Dr. B.S. Nagbhushan, Dr.Vijaya Kumar B.P., "*Investigations of issues and challenges in wireless sensor network for healthcare applications*", published in proceeding of National conference ETIT- 2010, September 17th & 18th 2010, pp 86-87, Bellary Institute of Technology & Management, Bellary.

2. Kiran Kumari Patil, Dr.B.S. Nagbhushan, "*A generic noninvasive approach for phonocardiography monitoring System in telemedicine application*", Proceedings of National Level Conference on Networking, Embedded and Wireless Systems (NEWS -2010), BMSE, Bangalore, Aug 2010, pp. 333-336.