Assisted Auscultation: Creation and Visualization of High Dimensional Feature Spaces for the Detection of Mitral Regurgitation

by

Daniel Demeny Leeds

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[June 2006]
May 2006

Author .....
Department of Electrical Engineering and Computer Science
May 26, 2006

Certified by ....
John V. Guttag
Professor, Computer Science and Engineering
Thesis Supervisor

Accepted by .....
.....
Smith
Chairman, Department Committee on Graduate Theses

Assisted Auscultation: Creation and Visualization of High Dimensional Feature Spaces for the Detection of Mitral Regurgitation

by

Daniel Demeny Leeds

Submitted to the Department of Electrical Engineering and Computer Science
on May 26, 2006, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Cardiac auscultation, listening to the heart using a stethoscope, often constitutes the first step in detection of common heart problems. Unfortunately, primary care physicians, who perform this initial screening, often lack the experience to correctly evaluate what they hear. False referrals are frequent, costing hundreds of dollars and hours of time for many patients.

We report on a system we have built to aid medical practitioners in diagnosing Mitral Regurgitation (MR) based on heart sounds. Our work builds on the "prototypical beat" introduced by Syed in [17] to extract two different feature sets characterizing systolic acoustic activity. One feature set is derived from current medical knowledge. The other is based on unsupervised learning of systolic shapes, using component analysis. Our system employs self-organizing maps (SOMs) to depict the distribution of patients in each feature space as labels within a two-dimensional colored grid. A user screens new patients by viewing their projections onto the SOM, and determining whether they are closer in space, and thus more similar, to patients with or without MR.

We evaluated our system on 46 patients. Using a combination of the two feature sets, SOM-based diagnosis classified patients with accuracy similar to that of a cardiologist.

Thesis Supervisor: John V. Guttag
Title: Professor, Computer Science and Engineering

Acknowledgments

I am greatly indebted to the many individuals who have provided me with great insights, guidance, and support throughout my work on this thesis.

My thanks go to my advisor, Professor John Guttag, who introduced me to the assisted auscultation project and has provided me with invaluable advice on my research and on my broader academic career. In my years of collaboration with him, Professor Guttag has taught me good research practice, has encouraged me to broaden my circle of collaborators, and has impressed upon me the importance of clear communication. The lessons he has taught me will continue to inspire and to direct me throughout my academic journeys.

My thanks also go to the many other scientists, both students and staff, who have offered useful insights. Zeeshan Syed and Dorothy Curtis dedicated significant time to orient me to the assisted auscultation system they had developed. As I took over the project, they offered feedback and new ideas. Also, I am particularly grateful to Ali Shoeb for pushing me to learn about and to apply self-organizing maps (SOMs), a tool that became central to my research.

Collaborating physicians helped me maintain my focus on the medical context of my work. Doctors Robert Levine and Collin Stultz both guided me with their knowledge of cardiac physiology and of auscultation practice. My thanks go to Doctor Levine for meeting with me to review specific details of my system. My thanks also go to Doctor/Professor Stultz for helping me organize a clinical study at the Veterans' Affairs Hospital and for listening to recorded heart sounds, enabling me to compare the performance of the assisted auscultation system to that of a cardiologist.

I owe much gratitude to the many friends who provided me with support and encouragement as I pursued my research and worked on this thesis. My thanks go to Jennifer Chiu for providing an honest, non-technical assessment of my use of SOMs; to Ashley Rothenberg for helping to push me through the writing process; and to David Chau for re-arranging his schedule to attend my Masterworks talk. Again, thanks to Ali Shoeb for his continual advice and good company in our Gates Tower

lab.

Finally, I owe my deepest love and appreciation to my family, which has guided my personal and intellectual growth throughout my life. At all times, they have celebrated my successes and helped my learn from my mistakes. Thanks go to my sister for providing her medical perspective on my use of SOMs for assisted diagnosis. Thanks go to my grandmother for inspiring me with the stories of her own years of research in America and in Hungary.

Contents

List of Figures

Chapter 1

Introduction

This thesis presents novel applications of artificial intelligence methods to aid primary care physicians in using heart sounds, heard through a stethoscope, to detect valvular disfunction. Building off the patient-specific "prototypical beat" introduced by [17], we assess the classification utility of several feature extraction techniques and explore a two-dimensional representation of our multi-dimensional feature spaces to provide intuition for the distribution of patients in these spaces.

1.1 Motivations

At a standard medical checkup, a physician uses a stethoscope to listen for signs of any problems in the patient's heart, in a process called "cardiac auscultation." Medical literature provides a wealth of information on the nature of ordinary and pathological (unhealthy) heart sounds. The frequencies and timing of heart beats can be indicators of the heart's health. For example, the presence of murmurs — whooshing sounds audible over portions of the beat—generally indicate some form of cardiac abnormality. [8]

Primary care physicians face several problems in interpreting the sounds they hear during auscultation. Doctors may not have time to experiment with the many ways of listening to the heart, since they must assess all elements of their patients' health. Also, doctors performing regular medical checkups do not have the time

to develop their ability to interpret heart sounds. Similarities between the acoustic characteristics of pathological and benign murmurs, murmurs that pose no risk to the patien, exacerbate this problem. Medical training also may mislead physicians, since standard recorded knowledge does not always conform to the observations of experienced cardiologists. [1]

As a result, physicians often incorrectly diagnose heart problems through auscultation. In [19], Vukanovic-Criley found the medical practitioners they tested missed 65% of systolic murmurs.[1]

A good non-invasive alternative to auscultation currently exists, but it is prohibitively expensive for use in ordinary checkups. The echocardiogram (or "echo") is treated as the gold standard for diagnosing a variety of conditions, including mitral regurgitation (the condition on which this thesis will focus). Sending high frequency sound waves into the body and observing reflections, the echocardiogram machine provides a picture of blood flow in the heart. This picture requires special training to interpret, but provides a definitive and (generally) accurate diagnosis. Unfortunately, the process costs three hundred to a thousand dollars per run, placing it out of the offices of primary care physicians. A high false referral rate, like the one anticipated by [19]'s statistic, results in a diversion of significant medical funds for unnecessary echos. The medical system could benefit significantly from an affordable means of identifying most falsely referred patients, or from an inexpensive replacement for the echocardiogram.

1.2 Goals

We seek to produce tools that will aid primary care physicians in their identification of patients with cardiac valve disfunctions, focusing on patients with mitral regurgitation (MR). While noting the difficulty in producing a perfect classifier based on solely acoustic features, we work towards allowing physicians to maximize the number of correct screening decisions they make based on sounds heard during auscultation.

[1] See Sections 1.3 and 2.1.1 for the definitions of "systole" and "diastole."

Because our system is directed towards primary care physicians, who would not formally diagnose a heart condition, it is important to note our system suggests only a screening decission. Positive classification, *i.e.*, the doctor thinks the patient might have a heart problem, will lead to a referral to a cardiologist for further evaluation. We seek to ensure a small percent of false negative classification (*i.e.*, the doctor will say "patient is healthy" when a patient has a pathological heart condition) and to limit false positives. We will judge the performance of our system against that of trained physicians listening to the same patient heart sounds, to establish more clear standards for our tools.

We want our aids to be inexpensive and straightforward to use, making them as accessible as possible. Presuming a doctor already has a computer in his/her office, s/he may run our software at no cost. Thus, our main monetary concern is that the heart signals required by our software be obtainable from devices already available to the primary care physician, or easily purchased. The outputs of our system should be easy to interpret by a doctor, or other medical practitioner, lacking technical background. We also aim to allow the user to form intuition about our software's operation, to have greater trust in its guidance and to detect if the system's suggestion for a particular patient is not reliable.

1.3 The Cardiac Cycle

We pause here to go over the basic structure of the cardiac cycle to better understand the assisted auscultation system introduced below. The heart consists of chambers that expand and contract periodically to circulate blood through the body. During "systole," two chambers called the left and right ventricles contract to push blood out of the heart. During "diastole," these chambers expand, accepting new blood into them. During auscultation, one can hear heart sounds, S1 and S2, daccompanying the beginning of systole and diastole respectively. [11] Section 2.1.1 provides much further detail about cardiac physiology.

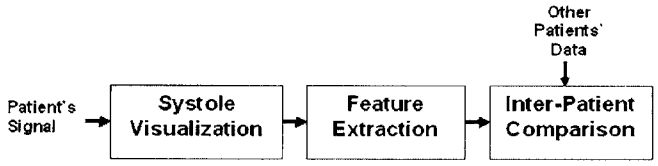


Figure 1-1: Diagram of automated auscultation system blocks.

1.4 Approach

We construct our diagnostic tool from three functional blocks working in series, as shown in Fig. 1-1: *systole visualization*, *feature extraction*, and *inter-patient comparison*. *Systole visualization* produces a visual summary for each patient's acoustic recording through a set of time-amplitude plots we call the "prototypical systole." *Feature extraction* measures a small number of feature values from the prototypical systole plots, attempting to capture its morphology. *Inter-patient comparison* presents each new patient in relation to already-diagnosed patients distributed across the selected feature space, producing either a visual comparison or a discrete classification.

Starting from the work of [17] and [13], we present modifications to each block, particularly focusing on new feature sets and on presenting the distribution of patients across these feature spaces. We compare our improved system against both the "gold standard" diagnoses of an echocardiogram and the ears of a cardiologist using only a stethoscope, to determine the difficulty of diagnosis based on our data set.

Our *systole visualization* follows the approach of [18], segmenting the original recording into individual beats and identifying the bounds of systole, splitting the signal in each beat into several "frequency" bands,[2] and, presenting the "average" shape of systole across beats. We introduce variations in the number of bands and scale all selected systoles to a fixed duration before determining the prototypical systole.

For *feature extraction*, we experiment with three different feature sets derived from

[2]In 3.1, we see that the bands do not correspond exactly to frequency ranges.

prototypical systole: generic, physiological, component-based.

We start with "generic features," which we produce by splitting each band of prototypical systole into one hundred bins and recording the energy in each one as a distinct feature. Viewing the feature values in succession is roughly equivalent to viewing a down-sampled version of the original prototypical beat, scaled to a have a constant duration and maximum height. Generic features are most useful as the basis for "physiological" and component-based feature sets.

Physiological features measure attributes of systole that medical wisdom has deemed useful in the diagnosis of MR. We search for three types of MR murmurs: holosystolic, "free standing," and "wide S2." Holosystolic murmurs have energy across all of systole. Free standing murmurs have energy isolated to some region of mid-systole. Wide S2 murmurs overlap with S2, making it appear unusually wide. We use five features, measuring energy, height, and offset of portions of the signal, in an attempt to capture any of the three types that may be present for a given patient.

Component-based features reflect comparison between the patient's prototypical systole and characteristic systolic murmur shapes discovered by principal component analysis (PCA). We rely on component/PCA-based and physiological features for the final system block.

For *Inter-patient comparison*, we use self-organizing maps (SOMs) to compress patient data from the high-dimensional feature space to a two-dimensional grid. Our system projects each newly-recorded patient onto the SOM grid already holding a set of diagnosed patients. We propose that the physician classify any new patient projected onto the grid based on the proximity of that patient's SOM data point to other points known to represent MR or lack of MR. If MR and lack of MR data points are close together in the map, the physician will know s/he cannot make a useful classification based on this data. In cases of doubt, or of positive diagnosis of the patient as having MR, the physician should refer the patient for further screening.

1.5 Contributions of this Work

This thesis presents an assisted auscultation system and discusses the potential utility of the techniques it employs.

- *Physiological feature set:* We show that this feature set provides a significant separation between most MR and non-MR patients Examination of individual features indicates most diagnostic content comes from a mid-systolic murmur detector defined in Section 3.2.2.

- *Component analysis:* We demonstrate that principal component analysis extracts recognizable morphologies from the training set of MR prototypical systoles. We also show that PCA-based features can help identify some patients with MR when physiological features are unable to distinguish these patients from those without MR.

- *Screening using self-organizing maps (SOMs):* We show how to classify patients based on their location in SOMs. We demonstrate how SOMs can be used to evaluate the classification utility of the underlying space, *e.g.*, by searching for regions with clear diagnostic significance on the grid and by comparing the prototypical systoles of patients placed together in the map.

- *Analysis of system performance:* Starting with a fixed set of acoustic recordings from patients with and without MR, we compare the diagnostic accuracy of a technician (the author) using our system with that of a trained cardiologist. We find our system allows a user to perform slightly worse than our cardiologist collaborator—a considerable achievement, given cardiologists tend to have significantly greater skill in auscultation than do other doctors, *e.g.*, primary care physicians, who typically screen patients. [19]

1.6 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 includes background on several elements of our research. We describe heart physiology and the factors involved in heart auscultation. We introduce self-organizing maps, providing the groundwork for inter-patient comparisons. Chapter 2 also establishes the background research in automated auscultation performed prior to the research within our group. We procede, in Chapter 3, with a detailed description of our automated system, addressing prototypical systole construction, feature extraction, and classification based on self-organizing maps. Chapter 4 follows with general methods used in testing and developing our system. We present and evaluate our findings in Chapter 5 and propose further work in Chapter 6.

Chapter 2

Background

This chapter presents medical and engineering background for our research. We discuss cardiac physiology and auscultation. We also review the mechanics of the two learning methods employed by our system.

2.1 Heart Sounds

Doctors have relied on heart sounds for over a century to reveal the state of their patients' hearts. Below, we discuss the physiological activities producing these sounds and the variety of techniques doctors use to detect different cardiac conditions.

2.1.1 Physiology of the Heart

The heart uses a combination of chambers and valves that move in a cyclical pattern to pump blood through the body.

The heart has two pairs of chambers, accompanied by two pairs of valves. The left and right atria receive blood into the heart and empty blood into their respective ventricles. The ventricles contract at regular intervals to push this blood out of the heart. Valves open and close to direct the blood pushed by the chambers. The aortic and pulmonary valves close to prevent blood outside the heart from flowing backwards into the ventricles while the ventricles are being filled by the atria. The mitral and

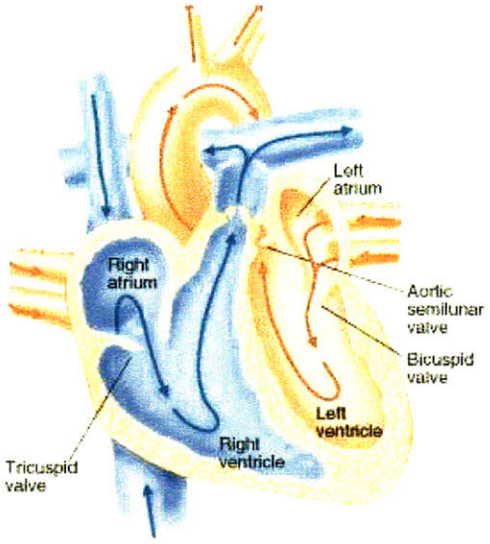


Figure 2-1: Chambers and valves of the heart. [3]

tricuspid valves close to ensure the contracting ventricles do not force blood back into the atria.

The heart cycle consists of two parts. During "diastole," the ventricles expand, or "dilate," to accept blood from their respective atria. The aortic and pulmonary valves are closed and the mitral and tricuspid valves are open to ensure appropriate blood flow. During "systole," the ventricles contract to push out blood, while the atria begin to fill with new blood from outside the heart. At this time, the mitral and tricuspid valves are closed while the other valves are open. Diastole usually lasts longer than does systole.

Our research focuses on detecting the failure of the mitral valve to remain closed throughout systole. In "mitral regurgitation," or MR, the mitral valve fails to close completely or opens backwards under the pressure exerted by the blood in the left ventricle. Thus, some blood is regurgitated into the left atrium, rather than continuing on its standard path into the rest of the body. [11] Many patients have a small degree of MR; physicians are concerned mostly with patients with unusually high MR—greater than or equal to 2 on the standard murmur intensity scale. [2]

2.1.2 Auscultation and Murmurs

The sounds audible through a stethoscope are influence by the closing of valves, the contraction of chambers, and the flow of blood. Normally, a doctor can discern the heart rate of a healthy patient and can distinguish between systole and diastole. The doctor often can maneuver the stethoscope and the patient to hear the abnormal heart sounds of a patient with a heart condition.

Most patient's heart sounda consist of two regularly-repeated thuds, known as S1 and S2. S1 and S2 each appear once, one after the other, in each heart beat. S1 corresponds to the closing of the tricuspid and mitral valves immediately preceding systole. S2 corresponds to the closing of the aortic and pulmonary valves at the end of systole. Thus, each "thud" actually consists of sounds coming from two distinct regions of the heart. Sometimes, the two events that constitute S2 occur sufficiently far apart in time to be distinguishable by ear. In [9], Durand describes ongoing controversy about the exact sources of S1 and S2, *e.g.*, the movement of blood or the contraction of muscles, and the frequency ranges of the resulting sounds. In this thesis, we ignore these finer physiological details.

Murmurs heard during auscultation indicate abnormal cardiac behavior, such as MR. Murmurs are prolonged acoustic activities characterized by their location in the cardiac cycle, their intensity, their frequency range, and their shape, or "morphology." An MR murmur, for example, will be high frequency and take place during systole, often near the middle or the end. The murmur is the result of turbulent flow of blood backwards through the partially-opened mitral valve into the left atrium.

A doctor determines the significance of a murmur in part based on the physical position of the patient and of the stethoscope during auscultation. Sounds originating from a given region of the heart radiate their acoustic energy as dictated by the orientation of that region and by the direction of turbulent fluid flow. Regurgitation at the mitral valve tends to radiate towards the apex and the axilla, whereas the sounds resulting from the stiffening of the aortic valve tend to radition towards the right upper-sternal border.

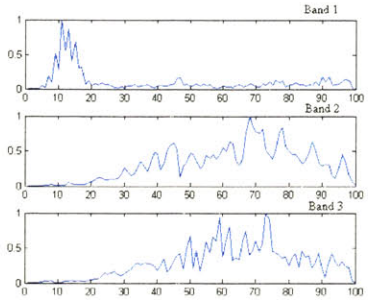


Figure 2-2: Sample output of systole visualization step.

Furthermore, a doctor can change the intensity of particular murmurs by having the patient squat, lie down, hold his/her breath, or take a vasoactive agent. Some techniques change blood volume levels; some change the physical orientation of the heart. If a murmur changes under different conditions, the doctor may be able to distinguish it from murmurs invariant across these conditions. [2, 8]

Auscultation has many challenges. There are too many combinations of patient and stethoscope positions to perform an exhaustive search for problematic heart sounds. Those sounds recorded at the "optimal" position for a given patient still may be too short or too soft to notice. MR murmurs often last less than a quarter of a second. Also, MR murmurs are attenuated as they travel through the chest wall, because the wall generally acts as a low-pass filter. The properties of this filter are not well understood, though [9] presents several time-dependent models.

Doctors' continued reliance on the stethoscope for nearly two centuries [9] indicates heart sounds have significant diagnostic content. New technologies may lessen some of the current challenges in auscultation, allowing doctors to gain more information from the acoustic signal.

2.2 Unsupervised Learning

We use two standard statistical methods to learn more about the relationship between heart sounds and diagnoses in our patient population. Principal component analysis attempts to model prototypical systole shapes for all our patients as the linear com-

bination of the shapes for a few fabricated "component" patients. The self-organizing map (SOM), in contrast, attempts to separate the population into several subsets, capturing the average properties of each subset and the inter-relationships among the groups.

Our learning methods work on the output of our systole visualization and feature extraction system blocks. The systole visualization block of our system condenses a sound recording from each patient into a short 3-channel "prototypical systole," such as the one shown in Fig. 2-2, representing typical acoustic energy during S1, systole, and S2. The feature extraction step translates each signal into a small number of features, as fully explained in Section 3.2.

2.2.1 Principal Component Analysis

Principal component analysis (PCA) models the set of MR prototypical systole shapes, X, as the combination of a set of principal components (PCs), P, ordered to provide the best linear approximations to X. Using this constraint, we expect the first PC, P_1, to show relatively high energy throughout systole, to capture the typical presence of high energy in the systoles of patients with MR. The next PCs could be added to the first to reflect the most common variation on the standard shape.

PCA is equivalent to the singular value decomposition (SVD) of the matrix

$$X = P\Sigma V^T. \tag{2.1}$$

To see the relation between SVD and PCA, we consider computing each PC in series. The first component, P_1, must capture the most energy possible:

$$P_1 = argmax_{||w||=1} E(w^T X)^2. \tag{2.2}$$

$w^T X_i$ will take on the highest value for a given X_i, the observed prototypical systole of patient i, when $w = X_i$. When there are multiple patients, we weight $w^T X_i$ by the probability of X_i occuring and compute the optimal w. The k^{th} PC is determined by

first subtracting the energy accounted for by the preceding $k-1$ components

$$\hat{X}_{k-1} = X - \sum_{i=1}^{k-1} P_i P_i^T X \tag{2.3}$$

and finding the PC accounting for the most energy in the remainder of the observed signals

$$P_k = argmax_{||w||=1} E(w^T \hat{x}_{k-1})^2, \tag{2.4}$$

bringing us back to the SVD form. [4, 12]

2.2.2 Self-Organizing Maps

A self-organizing map (SOM) represents its training data using a set of cluster centers. Each center reflects only the data points closest to it in the specified space, rather than attempting to model the whole population. However, the SOM provides a sense of the larger data distribution by maintaining an ordering among the centers and by recording distances between adjacent pairs. We use the SOM to find clusters of patients with similar feature values prior to the inter-patient comparison block.

The final location of centers is determined through an iterative algorithm. The SOM begins as a 1- or 2-dimensional grid of centers placed into the N-dimensional space in which the data lies; centers adjacent on the 1- or 2-dimensional representation of the grid are adjacent in the space. The grid may be initialized to a flat line or surface in the N-dimensional space, or to random positions. For any initialization, it is preferable the grid roughly spans the range of the data distribution. Next, we step through all the data points, adjusting the centers using the equation

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]. \tag{2.5}$$

For each data point $x(t)$, where $t = 0, 1, 2, ...$ is the iteration index, each center $m_i(t)$ is moved closer to the data point by a distance determined by the *neighborhood function* $h_{ci}(t)$. The neighborhood function decreases with increasing distance between $x(t)$ and $m_i(t)$, and with each new iteration t. One popular neighborhood function in the

literature [14] is

$$h_{ci}(t) = \alpha(t) \cdot exp(-\frac{||r_c - r_i||^2}{2\sigma^2(t)}), \tag{2.6}$$

where $\alpha(t)$ and $\sigma(t)$ decrease with t. The commonly-used distance metric is $||r_c - r_i||$. r_c is the location of the center closest to $x(t)$ and r_i is the location of the center being moved, *i.e.*, m_i. Thus, each data points pulls the closest centers even closer to it, while exerting a weak or non-existant force on distant centers. We repeat this algorithm until it converges or we reach a cutoff for t. [14]

The resulting SOM consists of many centers located in the middle of natural data clusters, and some centers spaced in un-populated space to preserve the original grid ordering.

There are many visualization tools to understand the location of centers and data properties each one represents. The "U-matrix" uses a grid of colored tiles to express the distances between center pairs in the original space, as well as the distances between the points associated with each center. Our Inter-patient comparison block relies on U-matrices. We can observe how many patients are assigned to each center. Using the terminology above, we determine the r_c for each x. Because many centers are assigned to specific data points, we can view these data points to derive a model for the local activity. In this thesis, we show that MR screening benefits from many of the properties of SOMs presented by [15].

2.3 Related Work

Over the past few decades, several papers have reported the ability to distinguish particular heart conditions from one another, or simply from regular heart operation, based on a variety of acoustical signal properties. Most studies of acoustic data affected by heart conditions have focused upon the characteristics of a few central heart sounds. There are many analyses of S1 and S2. Systole and diastole also have been examined. Some papers directly address the murmurs themselves.

In [16], Leung reports that temporal patterns in S2 can be used to distinguish S2 types reliably. S2 consists of two audible components, resulting from the closure

of two valves. Ordinarily, the aortic valve closes before the pulmonary, and the time between the two events changes with respiration. Unusual timing between the two parts of S2, *i.e.*, non-standard "splitting patterns," can indicate physiological abnormalities. [16] successfully tracks this timing data in sound recordings. This seems to be a promising technique for detecting conditions associated with "splitting patterns." [16] does not address MR and other conditions associated predominantly with systolic murmur rather than with S2 abnormalities. Also, [16]'s success with pediatric diagnoses may not carry over to work on adults, whose heart sounds are more difficult to hear.

Other papers have suggested more complex techniques. In [20], Zhang used the matching pursuit method to visually observe and compare murmurs. Splitting eleven sound recordings into time-frequency atoms, the authors reported that the utility of their method was highly variable based on the parameters chosen. Such instability could not be incorporated into a product for use by primary care physicians. Considering other options, the review article [6] mentions the employment of wavelets to extract murmurs from background noise. It describes using wavelets to better recognize the stenosis (narrowing) of arteries. An earlier paper by the same author [7] uses autoregressive methods to characterize diastole. While these results appear encouraging for the goal of recognizing coronary artery disease, the lack of further papers or commercialization of his approach in the fourteen years since publication may indicate that further investigation revealed faults in his work.

Most reports of progress towards a classifier for public use have appeared in conference papers. Researchers have reported as low as zero percent error from decisions based on the spectral and timing data of S1, S2, and murmurs. These encouraging results, however, may derive from a focus on children, who have clearer heart sounds, and by manual selection of data for analysis. Further discussion on the merits of these papers is available in [17].

Chapter 3

Techniques

Our assisted auscultation system combines a set of common techniques in artificial intelligence and signal processing with custom methods we develop for heart sound screening.

3.1 Defining "Prototypical" Systole

The systole visualization block condenses information from roughly 30 seconds of acoustic recording into a time-amplitude plot displaying a single "prototypical" systole, using the technique introduced by [17] and adjusted by [18].[1] Briefly, the algorithm takes in simultaneous acoustic and EKG recordings. It separates the acoustic signal into individual heart beats and extracts systole from each beat based on cues from the EKG signal. The algorithm then splits the time-magnitude acoustic signal in each systole into several frequency bands. Finally, a selected subset of systoles are "averaged" together to form the prototypical for each band.

Fig. 3-1 displays the constructed prototypical systole for a patient with moderate mitral regurgitation (MR). The peak of the Q wave (in the corresponding EKG signal, not shown) and the audible end of S2 are the designated limits of systole since they are relatively easy to locate in the original EKG and acoustic data. Because of

[1] In [18], Syed uses the term "prototypical beat" rather than "prototypical systole;" the remainder of this thesis uses the latter to acknowledge our focus on systole. We are optimistic that other studies will demonstrate the utility of a prototypical diastole.

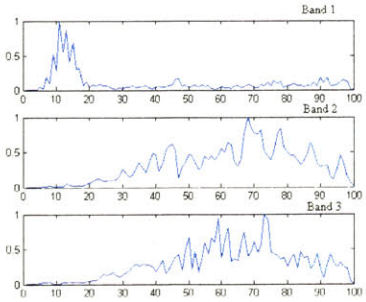


Figure 3-1: Sample prototypical systole.

the significant attenuation of high frequencies in heart sounds, as discussed in Section 2.1.2, we separate the prototypical signal into 3 bands, roughly corresponding to frequency ranges of 75-150 Hz, 150-300 Hz, and 300-600 Hz. Like [18], we apply wavelet bands to achieves this separation while providing high temporal resolution at high frequencies.

Prototypical systole incorporates data from the 3-band decompositions of a set of beats selected from the original acoustic signal. We introduce a processing step between the beat selection of [18] and the special "median" calculation for sample of systole described by [17], scaling each selected beat in time. Scaling employs vocoding to stretch all 3 band representations of systole to the duration of the longest one. By enforcing a standard length, timing characteristics are viewed as a percentage of systole rather than as a scalor offset from S1. This approach reflects the guidelines of medical knowledge that express the location of murmurs as "middle" or "late" systole, rather than as "300 miliseconds into systole" [8].

3.2 Features

Prototypical systole can provide the physician with a convenient visual overview of the acoustic signal for an individual patient. However, to uncover structure among groups of patients (which we can use for classification) our learning algorithms require that the tens of thousands of values making up each file's finely-sampled prototypical systole be summarized by a moderately small number of features.

This section presents several feature sets designed to capture the morphology of the signals produced in Section 3.1.

3.2.1 Generic Measurements of Signal Energy

Our "generic" feature set consists of a sequence of energy values measured at regular intervals between the start of S1 and the end of S2 in each wavelet band, mimicking the form of the original prototypical systole. In each band, we record energies from one hundred adjacent bins spanning prototypical systole and normalize by the maximum feature value.

Through normalization, we intend to account for the considerable inter-patient variation in acoustic signal amplitude. We posit that human listeners judge loudness of a given sound relative to the amplitudes of other sounds they hear in the acoustic environment. Unfortunately, normalizing based on a fixed activity (*e.g.*, S1 or the relatively fixed quiet of diastole, known as "diastolic floor"), presumes that fixed segments of the heart beat are constant across patients—a condition our experiments have shown to be false.

While prototypical systole can vary in duration, depending on the heart rate of the patient, the number of generic features stays fixed at one hundred for all patients. Thus, murmur onset and other timing data, when measured by bin index, effectively are expressed as percentages into systole, which are more informative in our problem than are scalar numbers of milliseconds from S1, as mentioned in Section 3.1.

Beyond the assumptions motivating scaling, the generic feature set does not reflect any knowledge about the characteristics of systolic murmurs. The other sets described below incorporate physiological and statistical models of pathological and normal heart sounds, benefiting from insights in cardiology and statistical learning, but potentially suffering from inaccuracies built into our models. Because the generic feature set, when plotted in sequence, roughly captures the original shape of prototypical systole in a fixed and relatively small number of samples—indeed, it acts as our visualization of prototypical systole in Fig. 3-1—it serves as the basis for these other sets. Below, we refer to the generic feature drawn from the n^{th} bin of band x

as $gen\_feat[x][n]$.

3.2.2 Extraction of Physiological Features

Following the approach of [13], this section identifies a small set of morphological measurements reflecting the properties of murmurs caused by mitral regurgitation (MR). We establish separate features for three types of murmurs:

- *Free standing:* When backflow into the aorta begins in the middle of systole, but tapers off to be inaudible prior to S2; in prototypical systole, it is not connected to either S1 or S2, and thus we refer to it as "free standing." Measure the height of the murmur and the temporal location of its peak.

- *Holosystolic:* When the mitral valve is unable to prevent backflow throughout systole, murmur may span from S1 to S2. Measure mid-systolic energy in each band.

- *"Wide S2":* When regurgitation begins in late systole and continues through the end of it, S2 and the murmur may overlap temporally, appearing as a "wide S2" in the prototypical systole. Measure energy and width of S2.

Focus on Band 3

Noting MR is associated with high pitched murmurs, as discussed in Section 2.1.2, we originally focused our search for murmurs on bands 2 and 3, which span from 150 to 600 Hz, and ignored band 1, which covers lower frequencies [18]. However, visual inspection indicated significant redundancy between the two higher frequency regions (*e.g.*, Fig. 3-1). Interested in minimizing redundancy, an approach known to improve performance of such classification techniques as support vector machines [10], we choose to focus on band 3 for the duration of this section.

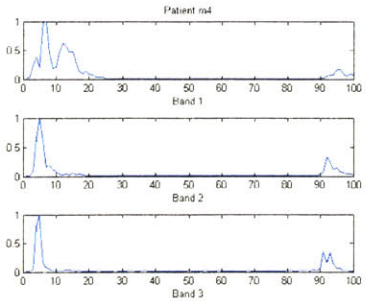


Figure 3-2: Prototypical systole for patient with MR.

Systolic Reference Measurements

For band 3 of the generic feature set, $gen\_feat[3]$, the starting time[2] of S2, as well as the median magnitude and variance of the systolic floor must be calculated before determining physiological feature values.

Knowing the end of S2 is roughly concurrent with $gen\_feat[3][100]$ for band 3, we begin our search for the start of S2 by finding the maximum value of S2, *i.e.*, the location of the maximum-valued $gen\_feat[3][n]$, constrained to $88 \leq n \leq 100$. Searching backwards from this point, called $s2\_peak\_loc[3]$, the starting location is the largest n such that $n < s2\_peak\_loc[3]$ and $gen\_feat[3][n]$ is sufficiently close to the systolic floor (as defined below) that we would not associate the recorded energy with the closing of the aortic and pulmonic valves. We label this point as $s2\_start[3]$.

To find the systolic floor in band 3, we divide the $gen\_feat[3][n]$ values in systole into four equal-duration sections and choose the section with the lowest energy[3]. We record the median and variance of $gen\_feat[3][n]$ values in this section, assigning them the variable names $syst\_floor\_median[3]$ and $syst\_floor\_var[3]$ respectively. Using these definitions, $n = s2\_start[3] - 1$ is the highest index below $S2\_peak\_loc[3]$ for which $gen\_feat[3][n] < (s2\_peak\_val[3] - syst\_floor\_median[3]) \cdot scale[3]$, where

[2]In this section, "time" signifies "percentage into prototypical systole," or, equivalently, "generic feature index number."

[3]Energy across a section is the summation of the individual $gen\_feat[3][n]$ values, as discussed below.

$s2\_peak\_val[3] = gen\_feat[3][s2\_peak\_loc[3]]$, and

$$scale[3] = \begin{cases} 0.3 & \text{if } syst\_floor\_var[3] > 5 \cdot 10^{-5} \\ 0.07 & \text{otherwise} \end{cases}$$

When systolic floor has moderate variance, portions of prototypical systole slightly above $syst\_floor\_median$ can be attributed to noise. However, when the floor is relatively smooth, slight bumps in prototypical systole are more likely to be physiologically significant. For example, a patient with a prosthetic tricuspid valve may have an unusually loud S1, as compared to the S1 produced by the closing of a natural tricuspid valve; a doctor, or patient, may hear the tricuspid valve close even without a stethoscope. Normalizing by S1's energy peak in each band of generic features will cause the substantially quieter S2, and a murmur of similar magnitude, to appear relatively quiet. However, they will be noticable with respect to the systolic floor, which looks flat after scaling. Fig. 3-2 illustrates this example. We use a conditional assignment rather than a standard mathematical function between $scale[3]$ and $syst\_floor\_var[3]$, such as $log(syst\_floor\_var[3])$, because of our empirical success with the former approach.

Free standing murmurs

To record free standing murmurs, we either must find the expected type of murmur, or, just as importantly, must produce useful output if such a murmur does not seem to exist. We begin by assuming the murmur is present, find it, and record relavent features. We then check our starting assumption and adjust features as needed.

Since we define these murmurs to be temporally separated from S2, and since at least some portion of MR normally occurs in the second half of systole, we declare the location of the murmur peak in band 3, $murmur\_peak\_loc[3]$, to be n such that $gen\_feat[3][n] = max_{50 \leq m \leq s2\_start[x]}\ gen\_feat[3][m]$ is the maximum value of those in the specified range. We also find the peak energy of the selected free standing peak

$$murmur\_peak\_height[3] = gen\_feat[3][murmur\_peak\_loc[3]] \tag{3.1}$$

A holosystolic feature, the mid-systolic energy, doubles as a third feature for free standing murmurs. *murmur_peak_height*[3] and *mid_systolic_energy* provide intuition about the height and width of the murmur, while *murmur_peak_loc*[3] provides knowledge about the murmur's relative position within systole.

We examine *murmur_peak_height*[3] to determine whether a free standing murmur actually exists. Specifically, we evaluate

$$murmur\_peak\_height[3] > scale[3] \cdot min(s1\_peak\_val[3], s2\_peak\_val[3]) \quad (3.2)$$

where $s1\_peak\_val[3]$ and $s2\_peak\_val[3]$ are the peak values of S1 and S2 in $gen_feat[3][n]$, and scale[3] is defined above. If our inequality is false, we assign $murmur\_peak\_height[3] = murmur\_peak\_loc[3] = -1$. We do not incorporate the systolic floor into our calculations in order to capture holosystolic murmurs, which will cause the systolic floor median to be close to the murmur peak value.

Holosystolic murmurs

As the energy of a holosystolic murmur spans from the end of S1 to the beginning of S2, the best feature set will capture the activity between these two points in prototypical systole. Unfortunately, it is difficult to record precisely within these boundaries. If high energy is visible throughout the band, standard approaches to finding the intended region of observation, *e.g.*, finding the proper "start" time of S2, often can fail. Expanding our view to record all the energy in each band ensures that all murmur activity will be captured. However, S1 and S2 widely vary in energy content across patients with and without MR, potentially obscuring the connection between a "band energy" feature and the presence of holosystolic murmurs.

To be able to capture most activity in these murmurs, when they are present, we fix two points in the middle of each band and sum the energy in the defined interval to find mid-systolic energy:

$$mid\_syst\_energy[3] = \sum_{n=25}^{75} gen\_feat[3][n] \quad (3.3)$$

The beginning and end points are chosen to exclude energy from most possible S1s and S2s, while including a sufficiently large portion of systole to record significant murmur energy, if it is present. Since each entry in $gen\_feat[x][n]$ is a measurement of the (normalized) energy in one bin of prototypical systole, the values need not be squared in Eqn. 3.3. As mentioned above, we consider mid-systolic energy to double as a feature for free standing murmurs; although we expect the measurement to reach beyond the temporal start and end of such a murmur (if we were to search for such markings), a more-careful measurement within these bounds would constitute a feature very similar to $mid\_syst\_energy[3]$, because the systolic floor will not contribute significant energy, thus running against our desire to avoid redundancy.

"Wide S2" murmurs

When the murmur resulting from MR overlaps with S2 in prototypical systole, the viewer can mistake murmur onset for S2 onset. We rely on a similar error in the calculation of $s2\_start[x]$ and use the resulting properties of the incorrectly marked S2 as "wide S2" features.

The appended murmur most prominently changes S2's apparent energy and width. Thus, we record:

$$s2\_energy[3] = \sum_{n=s2\_start[3]}^{100} gen\_feat[3][n] \tag{3.4}$$

and

$$s2\_width[3] = 100 - s2\_start[3] + 1 \tag{3.5}$$

Since prototypical systole ends at the termination of S2, we use 100 (the last sample of $gen\_feat[3][n]$) as the default maximum value of systole.

3.2.3 Feature Discovery through Component Analysis

While physiological features measure pre-determined morphological characteristics of three types of murmurs, principal component analysis (PCA) learns the most common murmur morphologies from a training set of generic feature bands. Our discovered

feature set consists of the projection of band 3 of gen_feat onto principal components.

PCA exploits a subset of the assumptions used in designing the physiological feature set. We determine all components from MR patients' gen_feat bands, presuming there is no substantial loss of information between prototypical systole and the generic feature set. Understanding that most, if not all, of murmur activity is confined to high frequencies, covered by the higher wavelet bands, we focus only on band 3. (We omit band 2 to prevent component analysis from focusing on inter-band variations rather than on inter-file variations.) Believing the variability of S1 can impair classification attempts, we ignore the first quarter of systole. Wanting to capture murmurs blending into S2, we keep the last quarter of systole in our data set. In sum, our frequency and timing specifications confine component analysis to $gen\_feat[3][n]$, for $25 \leq n \leq 100$.

Section 2.2.1 discusses the standard mechanics and statistical assumptions behind PCA. Working in Matlab®, we use the implementation developed by [5], respectively. `pcaproj` from [5] measures the similarity, via projection, between discovered PCA components and each patient's prototypical systole. Our component/PCA-based features consist of projection values for the first five principle components.

3.3 Self-Organizing Maps

As discussed in Section 2.2.2, U-matrices describing self-organizing maps (SOMs) allow a medical practitioner to gain intuition about the distribution of training recordings in feature space and about the position of an undiagnosed patient within this distribution. To create SOMs, we use the SOM Toolbox implemented by [5]. For a given training set, we normalize the data to be of unit variance across patients for each feature dimension and find a map in the feature space. Representing the map through the U-matrix, we place a label for each training and testing file on the file's best matching unit, *i.e.*, the map center closest to the file in Euclidean feature space. Labels for training files take the form `m#` or `#`. "#" is replaced by the file's index number; the label preceding the index indicates "**m**itral reguragitation," while the

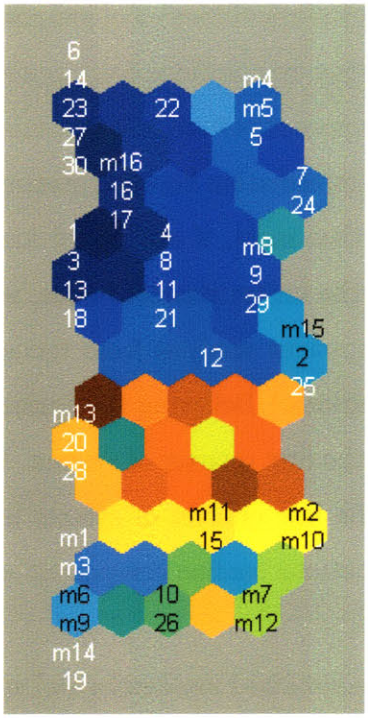


Figure 3-3: Sample self-organizing map.

lack of a label indicates "no mitral regurgitation." Labels for test files take the form `q#`. "q" indicates the status of the patient is in **q**uestion.

We can classify a `q` patient by projecting the him/her onto the map and comparing the location of the `q` label to that of labels for diagnosed patients. In Fig. 3-3, if `q1` appeared in the bottom left hexagon, we would label it "probably MR." If it appeared in the top left hexagon, we would label it "probably no MR." Section 5.1 discusses division of the SOM into diagnostic regions.

Chapter 4

Methods

In this thesis, we evaluate the outputs of our assisted auscultation system operating on a data set of 46 patients, split between those with and without mitral regurgitation (MR), and compare the classification performance of a user aided by this system with that of a cardiologist listening to the original acoustic recordings.

4.1 Data Collection/Selection

We use the data set collected for [17]. These files include auscultatory signals from the apex and from the parasternal notch, recorded from patients in the sitting, squatting, and lying-on-left-side positions. Each file also contains an EKG signal temporally aligned with the acoustic data; both signals stretch between thirty seconds and one minute. Over one hundred patients participated in the study (roughly thirty with mitral regurgitation and seventy without), each recorded between two and four times.

Each recorded patient also received an echocardiogram, which was analyzed by a technician to produce a "gold standard" diagnosis. For the purposes of our study, we label patients as having MR only if their echocardiogram reveals they have mitral regurgitation of an intensity greater than or equal to 2. Section 2.1.1 notes less intense murmurs generally do not merit medical attention. Throughout the thesis, we refer to patients "with MR" and "without MR" based on their echocardiograms.

We remove about half of the patients' records from our study following a few

guidelines. We wish to minimize extraneous sources of error from files a human (and, presumably, a computer) could easily recognize as un-informitive; therefore, we remove files with noisy EKG signals and with noisy or extremely low amplitude acoustic signals. We also wish to minimize the number of variables used in the composition of our training set; thus, we select files recorded from only one position on the body (the apex) with the patient in only one body position (lying on his/her left side). We choose these parameters as they maximize the number of files available for analysis and, from a physiological standpoint, constitute what tends to be one of the best positions for capturing the murmurs we seek, as discussed in Section 2.1.2. The selected files are split between 16 patients with mitral regurgitation and 30 without it.

4.2 Evaluating the Assisted Auscultation System

We run each patient through all of our three system blocks, studying the system's output. We first create a prototypical systole for each patient and extract the three features sets—generic, physiological, and component-based. Next we use SOMs to display the distribution of patient data in the spaces defined by the physiological and PCA-based feature sets, and study the results to determine the diagnostic value of each feature set.

4.2.1 SOM-Based Classification

We search for regions in each map with moderately clear diagnostic value, *e.g.*, several nearby hexagons predominantly hold patient with MR. We compare the prototypical systoles of patients placed close together on the maps, to determine if the SOM algorithm and the feature set successfully group patients based on their similarity to one another.

Evaluation of SOM "performance" is made difficult by the requirement of human judgement for any given diagnosis. Nonetheless, we employ standard leave-one-out cross-validation, incorporating human intuition into the testing loop. Given $n = 46$

files, we form a map from $n - 1$ files, project the training files, and then project the 1 test file. For each test file, we classify as "probably MR" or "probably no MR," based on the proximity of the patient's projected label to the labels in the SOM, as introduced in Section 3.3. If there appears to be insufficient information, we classify the file as "probably MR," indicating the patient, like other "probably MR" patients, should be sent for an echocardiogram examination to resolve our ambiguity. These choices are then compare to the echocardiogram diagnosis.

4.2.2 Comparison with a Cardiologist

We compare the classification guidance from our system, using both physiological and PCA-based features, with the classifications proposed by a trained cardiologist, Dr. Collin Stultz, on the same data set.

Because of his limited time, we had Dr. Stultz listen to half of our recordings—10 MR patients and 13 patients without MR. We selected most patients at random, although a few were intentionally included specifically because our system had difficulty with them.

We performed a double blind interview. Dr. Stultz listened through a set of headphones as we played files in a random order, determined by a computer random-number generator. For each patient, we asked him two yes/no questions:

- Does the patient have MR?
- Should the patient by sent for an echocardiogram?

We intended the latter question to permit him to express a degree of uncertainty, similar to the potential uncertainty one might draw from our SOMs. After a short time, we formally added the opportunity for Dr. Stultz to volunteer additional comments about each recording, noting he already had provided such comments without our prompting.

Because we intend our assisted auscultation system to help primary care physicians identify patients in need of further cardiac examination, we compare Dr. Stultz's

answer to our second question, "send for echocardiogram?" with the "probably MR"-"probably no MR" decision made by a technician (the author) guided by our system.

Chapter 5

Evaluation

Each of the three major system blocks contributes to the accessibility and the performance of our system.

In this chapter, we study the properties and performance of the three major system blocks: systole visualization (via prototypical systole), feature extraction, and inter-patient comparison (via self-organizing maps). Our modifications to prototypical systole formation help later blocks separate patients with and without MR. Physiological and component-based feature sets reveal distinct aspects of our patient population to be exploited during screening. Self-organizing maps (SOMs) presents non-technical users with a visual impression of patient data, as processed by the rest of the system. The SOM visualizations provide sufficient guidance for a user to distinguish between patients with and without MR with accuracy approaching that of a cardiologist.

5.1 Interpretation of Self-Organizing Maps

Our system's final output is a set of self-organizing maps (SOMs) intended to aid physicians in their screening decisions. Each map indicates pictorially the similarity between a new, undiagnosed patient, and several patients with known diagnoses, *i.e.*, "has MR" or "does not have MR." We equate subjective similarity of heart sounds with proximity of the corresponding data points in feature space, which is expressed by the SOM.

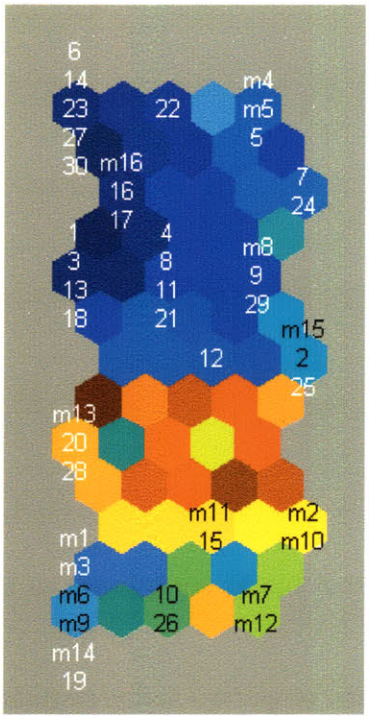


Figure 5-1: Sample self-organizing map, using physiological features.

In this section, we discuss how a user should interpret SOMs to make screening decisions. We employ these techniques throughout the chapter. We examine the merits of SOMs, and further SOM attributes, in Section 5.4.

Each patient in our data set appears once on each given SOM. We assign the patient a unique label, `m1` through `m16` for patients with MR and `1` through `30` for patients without MR. Undiagnosed patients receive a label of the form `q#`. Often in this thesis, we display the distribution of only diagnosed patients within the SOM, since we have no truly "undiagnosed" examples in our data set. We invite the reader to assess the ease or difficulty of classifying a recording based on its location in the map. In Section 6.2, we propose alternate labeling schemes when representing the data of hundreds or thousands of diagnosed patients. We believe the analysis methods proposed below for our 46-patient data set require little modification to be applied to our modified labeling approach for a larger set.

A user determines similarity between two patients based on their relative locations in the map and based on the colors of the hexagons in the vicinity of the patients.

A pair of labels located on opposite ends of the SOM identify patients whose heart sounds are quite different from each other. Map coloring indicates the degree of similarity between patients appearing relatively close to one another on the SOM. Patients sitting in a blue hexagon are more similar to one another than are those sitting on a red one.

We now review some elements of SOM construction to explain further the significance of the location and color of the hexagons in a SOM. The SOM algorithm adjusts a grid of points in the feature space to minimize the distance between patient data and the closest grid point; we call each point a "center." The hexagons in the SOM display represent centers. Between each pair of center hexagons is an "inter-center" hexagon whose color indicates the Euclidean distance in feature space between the centers on either side. The color of an inter-center hexagon indicates how similar the typical member of one adjacent center hexagon is to the typical member of another adjacent center hexagon. Because centers can be adjacent vertically, horizontally, or diagonally on the grid, all center hexagons not on the edges of the display are surrounded on all sides by inter-center hexagons. The color of each center hexagon indicates the distance between the corresponding center and the data points assigned to it (if there are any). For a hexagon holding patient labels, its color roughly indicates the similarity of these patients to one another.

A user can use the coloring of center and inter-center hexagons to identify broader diagnostic regions. For example, the predominance of blue hexagons, indicating high similarity, in the upper rows of Fig. 5-1 likely signifies patients sufficiently similar to one another to be considered one group. The ratio of patients without MR to those with MR in this broad region, 23 to 5, suggests that a doctor should label new patients appearing in the region as "likely not to have MR." The red and dark-red inter-center hexagons above `m13` and below `12` and `25` indicate patients in the bottom half of the SOM differ much more significantly from those in the top half than they do from one another, despite the lighter blue and yellow coloring in the bottom hexagons. Thus, we can declare the bottom portion of the map to be a second diagnostic region. The ratio of patients with MR to those without, 11 to 6, suggests new patients appearing

in the region are nearly twice as likely to have MR than not to have MR. We use the "60% heuristic:" if there are roughly twice as many patients with one diagnosis as those with the alternative diagnosis in a SOM region, *e.g.*, the bottom of Fig. 5-1, new patients in the region are labeled with the first diagnosis. Although 65% MR (or no MR) is a rather small majority, we find this practice produces relatively good system classification performance on our data set.

We should note we train on twice as many patients without MR as those with MR. Presuming our current patient distribution is representative, we can claim the true ratio of MR to no MR in the bottom region is 22 to 6, further supporting a doctor labeling new patients in this region as "probably MR." Unfortunately, doubling the number of MR patients also changes the ratio in the upper region to 23 non-MR patients to 10 MR ones, leading us again to rely on the 60% heuristic.

5.2 Prototypical Systole Formation

In [17], Syed introduces prototypical systole and addresses its value for medical diagnoses. Below, we evaluate the three changes we make to the original algorithm: time-scaling each systole before incorporating it into the prototypical; breaking the time-amplitude signal into wavelet bands (rather than [17]'s frequency bands); and selection of the most "representative" systoles through non-deterministic clustering. On the whole, each change is beneficial to our system or, at least, does no harm.

For a given patient, our systole visualization block splits the patient's acoustic recording into three wavelet bands, segments the signal into heart beats, and selects a set of "representative" systoles. Next, our system stretches the systoles to ensure all have equal duration. Finally, it finds the median amplitude value for each time sample across systole.

5.2.1 Time Scaling

Doctors typically judge murmur timing relative to the duration of the systole in which the murmur appears. To model medical practice, our system introduces time-scaling

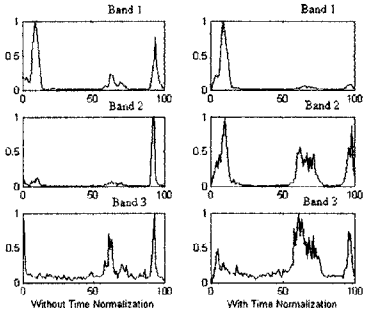


Figure 5-2: Prototypical systoles for patient m14, incorporating and omitting the time-scaling step. In this chapter's figures, "prototypical systole" will refer to the 300 generic features capturing the shape of the original prototypical systole.

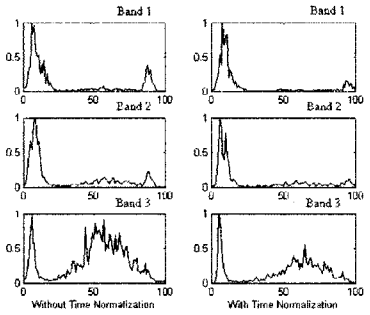


Figure 5-3: Prototypical systoles for patient m6, incorporating and omitting the time-scaling step.

into the formation of prototypical systole.

Time-scaling is advantageous to our general classification task, when compared to a system without time-scaling. This step typically allows prototypical systole to present a more clear view of any existing murmur morphology. It also leads to better separation between patients with and without MR in our feature spaces.

For individual patients, time-scaling reveals previously-hidden murmurs or keeps known murmurs visible. Fig. 5-2 demonstrates the significant growth of an MR patient's band 2 murmur using time-scaling. We suspect the patient has beats significantly varying in length, preventing the relatively-narrow murmur from lining up appropriately among the original systoles. Fig. 5-3 shows a decline in band 3 mid-systolic energy in MR patient m6. Nonetheless, the murmur remains clearly visible. Here, we suspect an abundance of short beats and of long beats allowed S2 of some

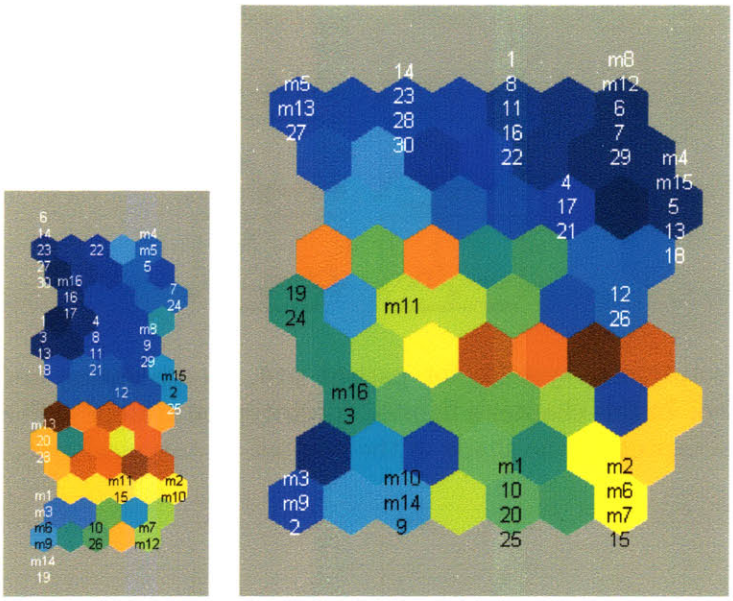


Figure 5-4: Self-organizing maps showing the distribution of patients in the physiological feature space based on their prototypical systoles, calculated (a) with and (b) without time-scaling.

short beats to be incorporated into band 3 morphology, accounting for the sudden jump in energy after a gradual rise from S1.

Our modification poses benefits to the physicians final screening decision, as he views the full population in our system's feature spaces. Fig. 5-4 compares self-organizing map (SOM) visualizations of the distribution of patients in the physiological feature spaces derived from the two types of prototypical beats.[1] In both maps, we see patients with MR tend to cluster near the bottom and patients without MR tend to cluster towards the top. However, only time-scaling produces a full red/orange line splitting the SOM into the "mostly MR" and "mostly no MR" regions. The high-distance, red separator produced without time-scaling fails to partition SOM (b) fully, indicating the "mostly MR" and "mostly no MR" regions are closer to one another in the physiological feature space (and more similar in appearance). A doctor should place more confidence in the separation and resulting diagnosis based on time-scaling.

[1]Later sections address our choice of features and provide full commentary on SOM analysis.

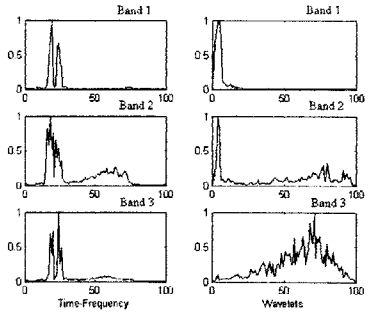


Figure 5-5: Prototypical systoles for patient m7, using filter bank and wavelet band approaches. The filter bank-based prototypical systole has low-magnitude tails at both ends, spanning roughly 10% of the prototypical distance on each side, corresponding to convolution beyond the edges of the original signal.

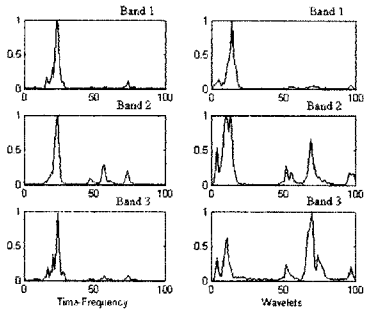


Figure 5-6: Prototypical systoles for patient m9, using filter bank and wavelet band approaches.

5.2.2 Wavelet Bands

We separate each patient's acoustic signal into three wavelet bands rather than using the frequency band approach of the original prototypical systole systole algorithm. In [18], Syed chooses wavelets to capture the short duration events at high frequencies, *i.e.*, MVP clicks. Our system applies the same technique to search for high-frequency energy spanning considerably longer periods of time.

Wavelet bands appear to strike a delicate balance between capturing high-frequency murmurs and generally missing high-frequency noise. Our technique usually heightens the visibility of MR murmurs in individuals' prototypical systoles, helping SOM-based screening.

Each wavelet band appears superior to it corresponding frequency band for our

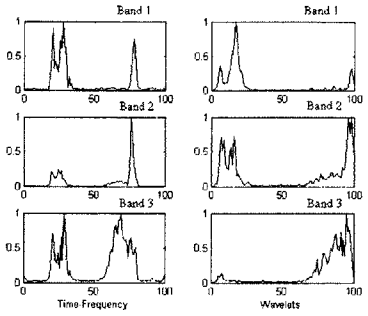


Figure 5-7: Prototypical systoles for patient m2, using filter bank and wavelet band approaches.

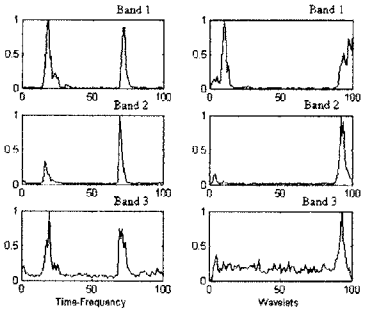


Figure 5-8: Prototypical systoles for patient 31, using filter bank and wavelet band approaches.

purposes because wavelet bands cover a broader range of frequencies. Although most energy in a wavelet band comes from a central range, nearby regions of the spectrum also contribute to the energy we observe. For example, band 3 focuses between 300 and 600 Hz. Frequencies slightly above 600 Hz and slightly below 300 Hz affect the band in proportion their distance from the central frequency.

Patient m7's prototypical systole, shown in Fig. 5-5, provides a good example of energy-blurring across frequencies. The shape appearing in wavelet band 3 is similar in form to the smaller murmur in frequency band 2, indicating energy coming from lower frequencies. Activity in higher frequency bands may help increase the visible energy in the shape.

Fortunately, blurring energy across frequencies generally does not capture enough irrelevant high-frequency noise to flood prototypical mid-systole with non-murmur-related energy. For most patients, wavelet bands feature an elevated systolic floor—a

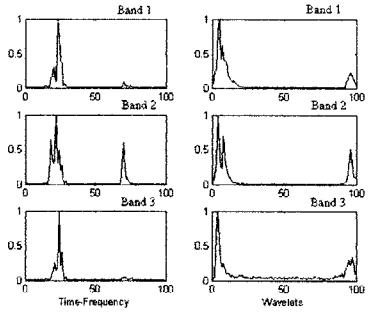


Figure 5-9: Prototypical systoles for patient 22, using filter bank and wavelet band approaches.

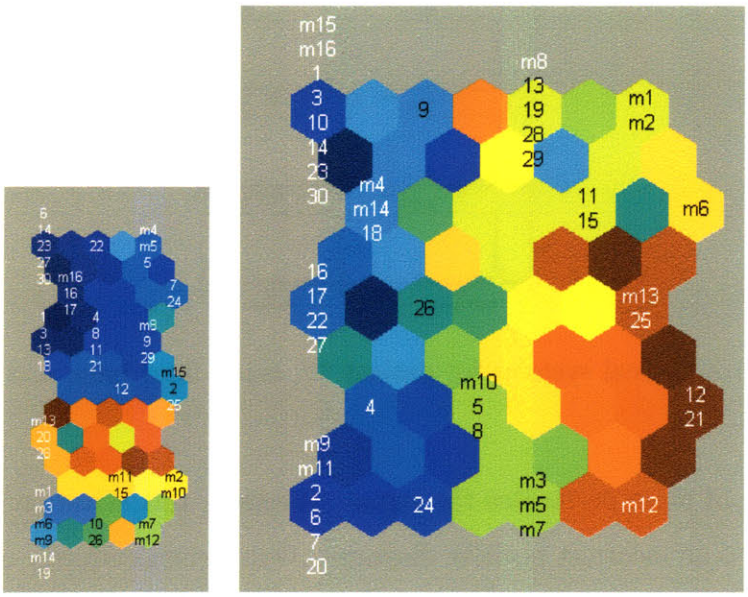
constant level of noise permeates the signal. However, this elevation usually is too small to lead to an incorrect positive diagnosis. Our system correctly marks patients 31 and 22, shown in Figs. 5-8 and 5-9, as "no MR" despite their non-zero systolic floors.

Comparison of SOMs resulting from the frequency band and wavelet band approaches, in Fig. 5-10, indicates wavelet bands may pose great benefit to patient classification. The frequency bands-derived feature space evenly spreads patient with and without MR throughout the map, providing little room to designate diagnostic regions. Discussion in Section 5.1 already has established these regions in the wavelets-based map. The striking visual difference between SOMs seems sufficient motivation to favor wavelet bands above frequency bands.

5.2.3 Non-Deterministic Clustering

The prototypical systole is formed based on a subset of all systoles in a patient's acoustic recording. This subset is selected by assigning each systole two feature values, as described by [18], clustering the systoles, and choosing the largest cluster. The selected subset may vary in content from run to run, due to the non-deterministic nature of clustering.

We observe our system may assign as few as 8 heart beats to the largest cluster. This figure is reasonable in order of magnitude, given we typically capture about 30 heart beats in 30 seconds of recording. We cannot increase significantly the number



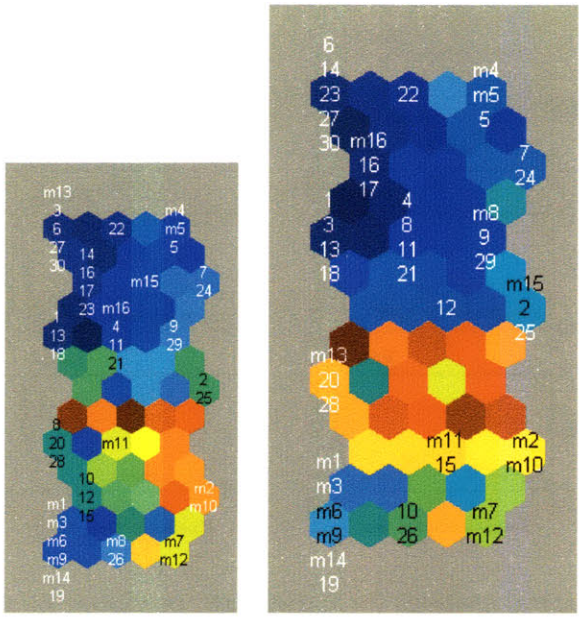
(a) (b)

Figure 5-10: SOMs showing the distribution of patients in the physiological feature space based on their prototypical systoles, calculated with (a) wavelet bands and (b) frequency bands.

of selected systoles without recording for several minutes—an infeasible option for the primary care physician intended to use our software. In [17], Syed also establishes that his system is biased to select the longest beats, most likely to reveal MR. The length of beats follows the breathing cycle, which is usually much slower than the cardiac cycle, leading to the inclusion of less than 30% of recorded beats.

Because prototypical systole formation relies on a low-population cluster, we are concerned that prototypical appearance might change considerably as our non-deterministic clustering algorithm changes cluster members on each run. However, [18] establishes performance benefits gained by the clustering step.

Fortunately, our experiments indicate that prototypical systoles, and their resulting distributions across the feature spaces, remain relatively stable across runs of our system. Once again using SOMs to visualize the distribution of patients in physiological feature space, derived from two copies of identical prototypical systole formation code, we find relatively minor differences. Fig. 5-11 shows the "mostly MR" and "mostly no MR" regions are retained with the red, high-distance separator across the



(a) (b)

Figure 5-11: SOMs showing the distribution of patients in the physiological feature space based on their prototypical systoles using two sets of prototypical beats created with identical time-scaling and wavelet band settings. We rely on the second run, (b), for the rest of the thesis.

two maps. The distribution of patients within each region also remains similar, with many patients falling into the same hexagon in each SOM.

Although a small number of patients, like MR patient `m8`, move between distant hexagons from one map to the other, further inspection shows these changes may be insignificant to classification. `m8` moves from the bottom in Fig. 5-11(a) to the upper-right region in Fig. 5-11(b). However, analysis of the second SOM in Section 5.4 suggests this right edge be considered an MR block within the "mostly no MR" region. Visual comparison of the two underlying prototypical systoles in Fig. 5-12 shows relatively little visual difference, suggesting a weakness in the physiological features for classifying this patient rather than a problem with clustering.

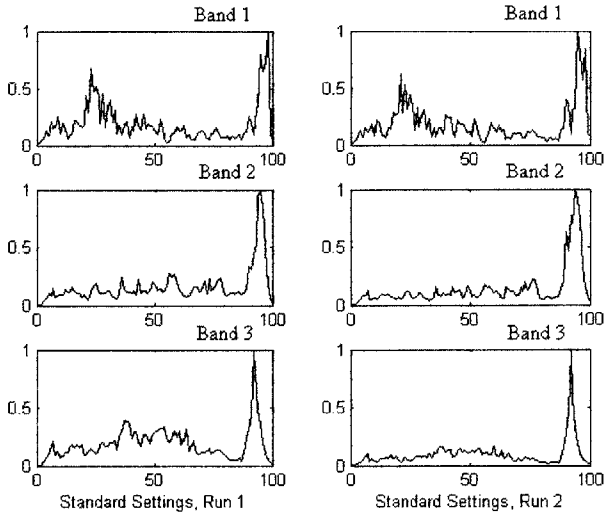


Figure 5-12: Prototypical systoles for patient m8, produced by two runs of our final code.

5.3 Features

Section 3.2 describes the extraction of 3 different kinds of feature sets from prototypical systole: generic, physiological, and component-based/PCA-based. Experiments indicate each kind of feature set is useful to MR screening, although generic features provide significantly less help than do the other features we tested. We find projection of our patients into either physiological and PCA-based feature spaces successfully separates most patients with MR from most of those without MR. The SOM classification approach in each of these two spaces leads us to correctly identify 14 out of 16 MR patients. In contrast, SOM visualization of our patient distribution in the generic feature space shows moderate mixing of patients with and without MR, making a screening decision difficult.

5.3.1 Generic Features

To measure generic features, we divide each wavelet band of prototypical systole into 100 bins of equal size and record the signal energy in each bin. In developing these features, we intended them as an aid to the construction of other feature sets, rather than to facilitate direct inter-patient comparisons. Unsurprisingly, they prove

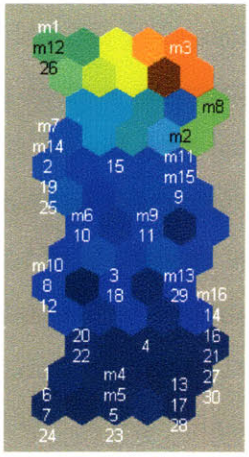


Figure 5-13: SOM showing the distribution of patients in the generic feature space based.

sub-optimal for classification tasks.

The generic feature set has many characteristics considered undesirable when training and running classifiers, including high dimensionality and significant redundancy among features. Despite these limitations, we observe generic features allow us to separate some patients into diagnostically-meaningful classes using SOMs. Screening with other feature sets yields still more reliable and more convincing diagnoses, as discussed in later sections.

Perhaps the most troublesome characteristic of the generic feature set, from the perspective of machine learning, is its size compared to the number of patients being studied. We measure three hundred feature values for each patient, as described in Section 3.2.1. Using our selected recordings, we would train on 46 patients. Thus, the distribution of data points in the feature space will be far too sparse to draw significant conclusions. Performing unsupervised learning, using SOM inter-patient comparison in Fig. 5-13, we see a relatively uniform coloring, indicating most patients are relatively evenly distributed across the feature space, showing no notably distinct clusters.

The high dimensionality of our feature set presents a second problem: it will present a greater challenge to users attempting to understanding the meaning of a classifier's output, or of a SOM visualization. Because of human's difficulty with

reasoning in more than 3 dimensions, they probably will have trouble visualizing the slope of the high dimensional separation surface produced by the SVM. Studying the prototypical systoles associated with the support vectors provides some insight into the surface, but it will prove intractible to study the hundreds of vectors chosen to capture the distribution of the thousands of training points. Undertanding the distribution of SOM centers, an analysis technique discussed in detail in Section 5.4, also will prove intractable given the number of training points. The generic feature space is not amenable to human comprehension of our system.

Our large feature set also may suffer from redundancy among features, which can be harmful to standard classification approaches. It is likely that many feature values are strongly correlated with one another, since they often are taken from adjacent or near-adjacent stretches of prototypical systole. Guided by [10], we suspect this redundancy contributes to poor SVM performance. We design later feature sets in part to avoid this problem.

Despite our concerns about using the generic feature set to separate our small patient group into diagnostic classes, we observe our SOM analysis approach identifies a "mostly MR" and "mostly no MR" regions in the generic feature space, in addition to a moderately large region of uncertainty. We might place up to nine MR patients into the "mostly MR" class at the top of the SOM. Unfortunately, both diagnostic regions blend into the "unclear classification" region—there is no notable separation between classification groups in space. Nonetheless, the presence of these regions might correspond to an existing separation in general feature space that could be revealed with more training data.

5.3.2 Physiological Features

Building off the generic feature set described in Section 5.3.1, we find physiological features to be useful in screening for MR. Thresholds on each of several individual features separate the patient population moderately well between those with and without MR; each feature has a distinct physiological significance, making the thresholds meaningful to a medically trained user. Projecting the patients into the full feature

space enables us to identify correctly more of those with MR, though sometimes at the expense of incorrectly labeling more patients without MR. Viewing patient distribution via SOM, we see the small, five-dimensional, space produces distinct diagnostic groups, unlike the generic feature space. In combination with results from screening in the PCA-based feature space, physiological features enable our final assisted auscultation system to miss only 2 out of 16 MR patients, as discussed in Section 5.4

Individual Features

We recall from Section 3.2.2 that the 5 physiological features in our set are (1) the height of the free standing murmur peak, (2) the offset of the free standing murmur peak into systole, (3) the mid-systolic energy, (4) the width of S2, and (5) the offset of the beginning of S2 ($s2\_start[3]$)). For this section, we often refer to these features by their number as listed here.

Two of the physiological features have significant classification utility when considered individually. Each of the two free standing murmur features can be used individually to perform threshold classification with accuracy close to that of screening with the full set using a SOM.

The first two plots in Fig. 5-14 show the distributions for the two free standing murmur features (features 1 and 2). By visual inspection, we can place a threshold line at $featurevalue = 0$ in either plot, separating the majority of patients with MR (above the line) from those without (below the line). This technique produces 5 false negative classifications (out of 16 MR patients) and 6 false positives (out of 30 patients without MR), equivalent to our results from SOM-based classification in the full physiological feature space.

Our results indicate the rough free standing murmur detector implemented in Section 3.2.2 may be a sufficient classifier on its own. If our system does not find a free standing murmur, it outputs the first two feature values as -1, below the threshold. If we do detect a murmur, we output a positive value that will be above our visually-selected line. Our simple threshold places no restrictions on the activity's temporal properties or energy content to be accepted as MR. Also, because a patient can only

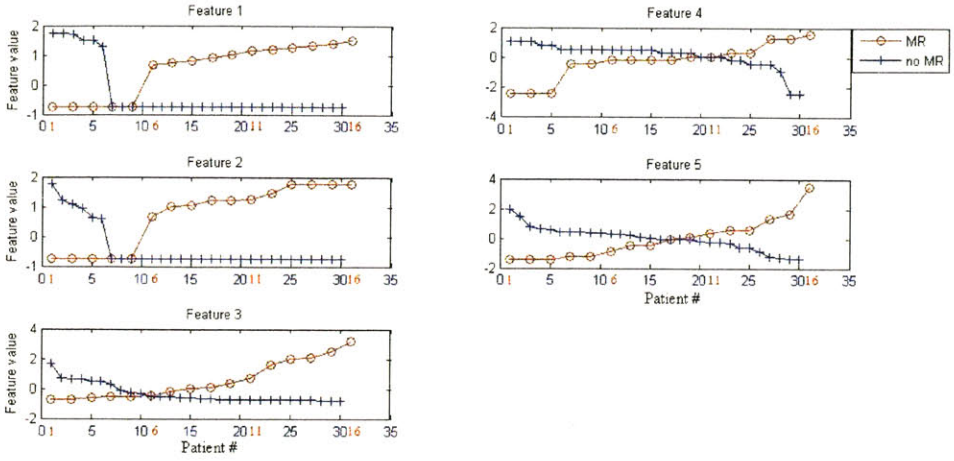


Figure 5-14: Individual feature plots. Each plot shows the values for all patients drawn from one dimension in the feature space, *e.g.*, mid-systolic energy. For easier comprehension, MR patient values are sorted to be in increasing order and non-MR patient values are sorted to be in decreasing order. For comparison, MR data points are stretched over the x-axis to occupy the same range as data point for patients without MR.

have a positive value for his/her first two features if we detect a free standing murmur, the threshold in either of the first two features misclassifies the same patients with and without MR.

The free standing murmur detector operates like the human who listens to the heart sound or who views our system's prototypical systole display. The presense of mid-systolic high frequency energy of sufficiently high magnitude makes us suspect a murmur. Holosystolic and "wide S2" murmurs share this characteristic with free standing murmurs, leading all three types of MR outlined in Section 3.2.2 to be captured by the simple detector. Perhaps the offset and height may distinguish between sub-classes, a task outside the stated goals of our system.

The last three features shown in Fig. 5-14 do not appear to be useful individually for classification, nor do they appear to identify MR patients misclassified by free standing murmur features. Although the optimal threshold on mid-systolic energy (feature 3) achieves the same number of false negatives as do the free standing murmurs, false positive classifications increase. Placing a second threshold, we can

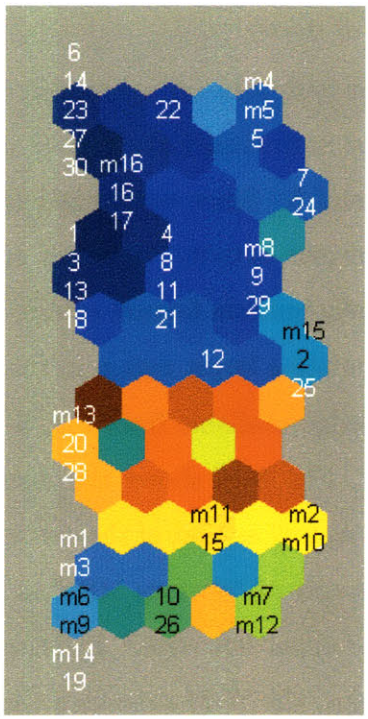


Figure 5-15: SOM distribution for full data set in physiological feature space.

isolate 4 further "MR" patients, definitively above the highest "no MR" patients; however, it is unclear whether further data from both patient groups would support this separator. "Wide S2" features (features 4 and 5) exhibit even less diagnostically helpful structure. The "wide S2" murmurs, which successfully are identified by these features (they are assigned the highest feature 5 values), already are identified by the free standing murmur features, as anticipated above.

SOM Display of Patient Distribution

In Fig. 5-15, we see that the SOM visualization of our patients' distribution in the physiological feature space, reveals even further structure. The SOM shows a clear separation—represented by two lines of red hexagons—between the majority of MR patients and the majority of those without MR. We observe an identical separation in the distribution of values for the first physiological feature, shown in Fig. 5-14; patients who fall below 0 in the one-dimensional plot fall below the upper red line in the SOM. Further features contribute to grid coloring, as well as to further spatial

separation: the lighter blue of the MR region, compared to the nMR region, indicates patients with relatively clear MR take on a broader range of feature values than do patients without MR. The individual feature plots in Fig. 5-14 demonstrate this phenomenon for features 3 and 5. Since we defined our measurements roughly to correspond to degrees of murmur activity, it is not surprising that most patients lacking murmurs have relatively uniform feature values, while those with murmurs of varying shapes and sizes populate a larger region in the feature space.

Besides presenting knowledge available from Fig. 5-14 and SVM tests, albeit in a compact graphical format, Fig. 5-15 shows the distribution of the SVM's false positives and false negatives, as well as that of the correctly classified files, within the MR and nMR blocks. We find a core of "obvious MRs" in the bottom left hexagon and of "obvious nMRs" along the upper left edge.[2] Ignoring the one outsider in each of the two aforementioned groups, the SOM organizes the set of prototypical systoles to indicate two prominent morphology groups. We may focus on such patients in training doctors to ensure they catch the most easily detected murmurs. Further studying the nMR block reveals that almost all patients with MR lie on the right edge of the SOM, and are packed less tightly in the feature space--a phenomenon that may prove useful to classification.

5.3.3 Component Analysis-Based Features

Component-based features use projection to measure the degrees of similarity between a patient's prototypical systole and automatically-derived components. We consider components created by principal component analysis (PCA), a statistical method designed to find standard morphologies from a training set of MR patients' prototypical band 3s. We only view the last three quarters of band 3, ignoring S1 activity.

We had only moderate success using component analysis and component-based features to characterize and to classify the prototypical systoles drawn from our data set. PCA extracts shapes that individually represent commonly-seen patient mor-

[2]The dark blue spreading across the upper left edge indicates the centers for each hexagon are quite close to one another in Euclidean space; thus, we can treat the edge here as one hexagon.

phologies, *e.g.*, typical forms of MR murmurs and of S2. We use these shapes to create the component-based feature set.

The PCA-based feature space is valuable in screening for MR, particularly when used to supplement other spaces. Compared to the physiological feature space, PCA space provides less clear separation between patients with and without MR. However, it allows a user to identify correctly several MR patients embedded in the "mostly no MR" region of the physiological feature space. In Section 5.4, we combine classification results from the PCA-based and physiological feature spaces to correctly classify almost all MR patients.

PCA-Derived Components

We use software from [5] to perform PCA on the prototypical systoles of our MR patients. Earlier principal components account for more of the energy—here, have higher corresponding eigenvalues—than do later ones. Because of details in the algorithms used to perform PCA, some discovered components are dominated by negative peaks, although only positive energy peaks have valid physiological interpretations. We flip these components, *i.e.*, scale by a negative number, when using them to reconstruct prototypicals in the data set

Because PCA attempts to capture the maximum energy across all patients, it focuses on low frequency shapes in our prototypicals. Individual principal components (PCs) capture long-range activities representing murmurs, S2, or both. We also can reconstruct common patient morphologies using relatively few components.

Some PCs in Fig. 5-16 have particularly clear relations to diagnostically-relavant attributes of systole, once we account for typical side-effects of PCA. The first PC, when multiplied by a negative scalar as discussed above, takes the form of a strong holosystolic murmur with a muted S2. In Fig. 5-17, we see one of our MR patients, `m6`, with a murmur rising and falling across systole, similar to the behavior of the inverted PC. Some PCs include energy over a broader range than seen in any patient; PCA creates components with extra peaks, in the opposite direction from the main peak of the PC, to cancel out extra energy from these wide components. Ignoring

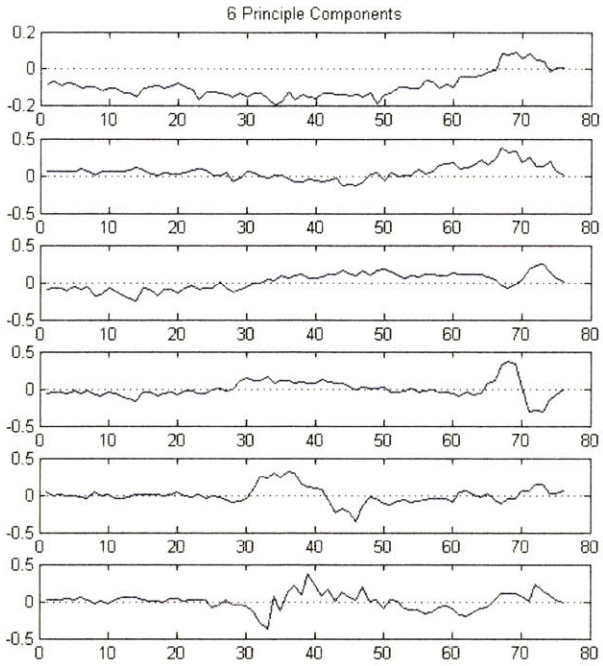


Figure 5-16: Six extracted principal components.

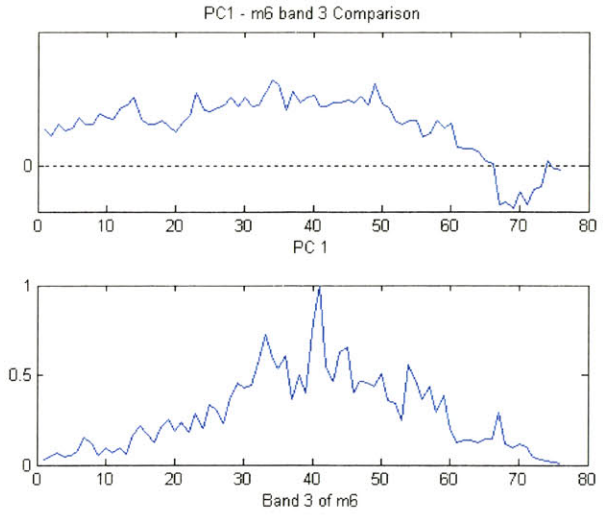


Figure 5-17: Comparing first principal component (PC) to last 75% of band 3 prototypical systole for patient `m6`.

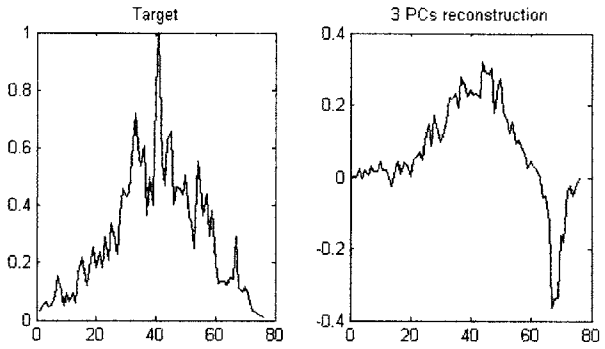


Figure 5-18: Reconstruction of m6 using three principal components.

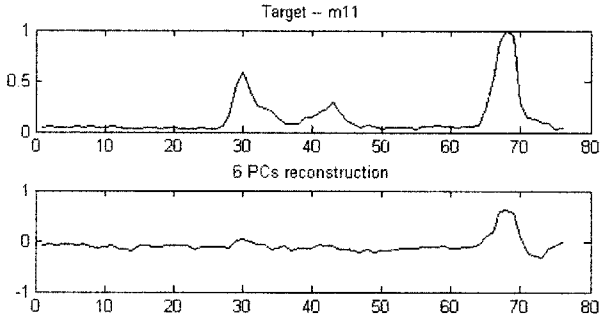


Figure 5-19: Reconstruction of m11 using principal components.

the negative peaks, we associate PCs 4 and 5 respectively with quiet MR, in which S2 is much louder, and with a strong MR murmur isolated to mid-systole. The other components relate to further S2 and MR murmur shapes, although the scaling of our display can make the activities difficult to see.

Not surprisingly, principal components are very useful for reconstruction of band 3s observed in our data set. The first 3 PCs can form an overall shape quite similar to that of the holosystolic murmur of m6, as shown by Fig. 5-18, ignoring the negative peak in the S2 region. Unfortunately, some band 3 murmurs cannot be reproduced be our components. The free standing murmur of m11, shown as the target in Fig. 5-19, does not match well with reconstruction from the first 6 PCs. PCs, in particular, fail to mimic the sharp peaks at samples 30 and 45 in the target.

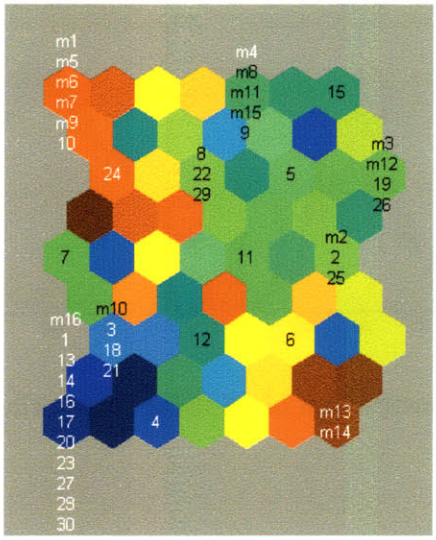


Figure 5-20: SOM distribution for full data set in PCA-based feature space.

PCA-Based Features

We create a new, five-dimensional feature set by projecting prototypical systoles onto our top five principal components.

PCA-based features make an important contribution to the MR screening system, although they perform worse than do physiological features. SOM visualization of the patient distribution establishes "mostly MR" regions holding MR patients hard to classify correctly in physiological feature space.

Traditional classification via simple thresholds performs unacceptably poorly in PCA-based feature space. The distribution of patients over any individual feature appears to be devoid of useful structure. We find no feature comparable to mid-systolic murmur peak height in Section 5.3.2, that neatly divides our population.

The SOM display of the patient distribution in PCA-based feature space, shown in Fig. 5-20, can help a user correctly classify MR patients misdiagnosed with physiological features. Patients with MR lie at the edges of the map most distant from the bottom-left corner, and nearly 50% of patients without MR are focused in this corner. We easily can define "mostly MR" and "mostly no MR" regions, although the blur of colors across the map informs us these regions do not consist of data points densely concentrated in the feature space. Most encouragingly, we find all but one of

the patients typically misclassified in the physiological feature space, *i.e.*, `m4`, `m5`, `m7`, `m8`, and `m15`, joined together in a "mostly MR" region at the top of Fig. 5-20.

5.4 Properties of Self-Organizing Maps

Throughout this chapter, we have used self-organizing maps to reveal structure in the high-dimensional distribution of our patient population. Below, we discuss tests of our SOM-based classification system, and the utility of the SOM visualization.

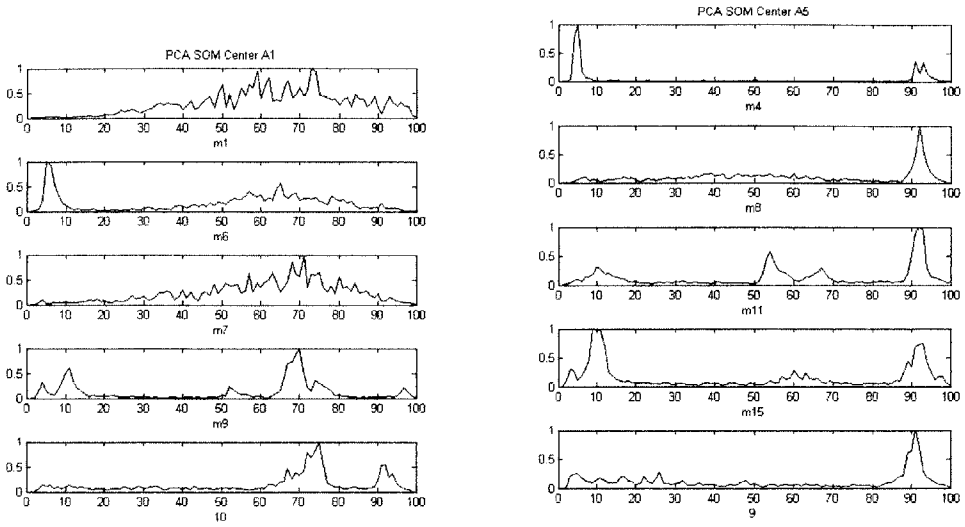
Our SOMs present non-technical users with screening recommendations with varying degrees of certainty. Our system projects most new patients into a "mostly MR" or "mostly no MR" region, suggesting a classification decision. Hexagonal coloring and the morphological display of neighboring patients' prototypical systoles informs users how much confidence to place in their decisions. Users also can use SOM appearance to determine which set of features (or, possibly, of patients) will best help them screen a given new patient. In a field where incorrect diagnoses can have serious consequences and where qualitative statements abound, our new classification system invites regular examination of its own performance, rather than requiring the user to rely upon the binary answer of a black box.

5.4.1 Distribution of Full Patient Data Set

SOMs display the distribution of patients in our high-dimensional feature spaces in a two-dimensional grid. These maps provide both a population-wide and fine-grained view of the distribution. Pursuing the fine-grained details, we can determine our trust in the feature space underlying a given SOM by comparing the prototypical systoles placed into the same hexagon.

Observing the PCA-Based Feature Space

The benefits of SOM visualization for classification and for more detailed data analysis are most clear in the PCA-based feature space, displayed in Fig. 5-20. The spatial orientation of labels leads us to recognize "mostly MR" and "mostly no MR" regions.



(a) (b)

Figure 5-21: Band 3 morphologies for patients in centers (a) "A1" and (b) "A3" of the PCA-based SOM shown in Fig. 5-20. We label centers by row and column; each row is assigned a letter (starting with A at the top) and each column is assigned a number (starting with 1 at the left).

The distribution of colors in the map helps explain why a radial basis kernel support vector machine running in the same space will perform worse than SOM-based classification, misclassifying 8 out of 16 patients with MR in leave-one-out testing. MR labels are located in red and green hexagons, indicating they appear sparsely in their section of the feature space. In both the upper-left corner of the SOM and on the right end of the second row of center hexagons,[3] "no MR" labels appear in MR-dominated space. A radial basis kernel will not recognize these labels as outliers; instead, it will over-fite and carve out unwarranted islands of "no MR" classification territory.

Understanding the Location of SOM Centers in Feature Space

Viewing band 3s of the prototypical systoles assigned to a given center gives us further intuition about each map's organization.

[3]See Section 5.1 for a definition of "center hexagons."

Inspection of band 3 of the prototypical systoles of patients appearing in individual centers in Fig. 5-20 indicates PCA-based features detect morphological similarities. Many centers hold a set of visually similar morphologies. Patients in the upper-left hexagon, seen in Fig. 5-21(a), all show clear murmurs reaching a peak around the 70% mark. Patients in other centers appear to have greater variability, but remain in the same diagnostic class. The populated center in the upper-middle of the SOM captures low magnitude murmurs of varying shapes, seen in Fig. 5-21(b); it correctly includes `m4`, a murmur with activity so small compared to the S1 peak that it is practically invisible. We appreciate that PCA-based features, unlike physiological features, stress form over energy.

Examination of murmur shapes for patients grouped together in physiological feature space provides similar confidence in Fig. 5-15, as well as intuition into the morphological significance of physiological features. The shapes in Figs. 5-22 and 5-22(a) illustrate the SOM's "recognition" that there is a difference between wide S2 murmurs and others, *e.g.*, holosystolic and free standing murmurs. The SOM draws these two subgroups respectively to the top-right and upper-left of the "mostly MR" region. Comparing the upper-left and upper-right corners of the "mostly no MR" region, in Figs. 5-22(b) and 5-22(c), we similarly see the expected difference between "clear no MRs," lacking any activity, and the higher-activity right edge "MRs and no MRs," grouped together because of slight turbulence in systole (often coincident with MR).

5.4.2 Pseudo-Classification

Leave-one-out testing shows that the SOMs for the physiological and PCA-based feature spaces are useful for classifying patients in our data set. Consulting both feature spaces for classification labels leads to detection of nearly all of our MR patients.

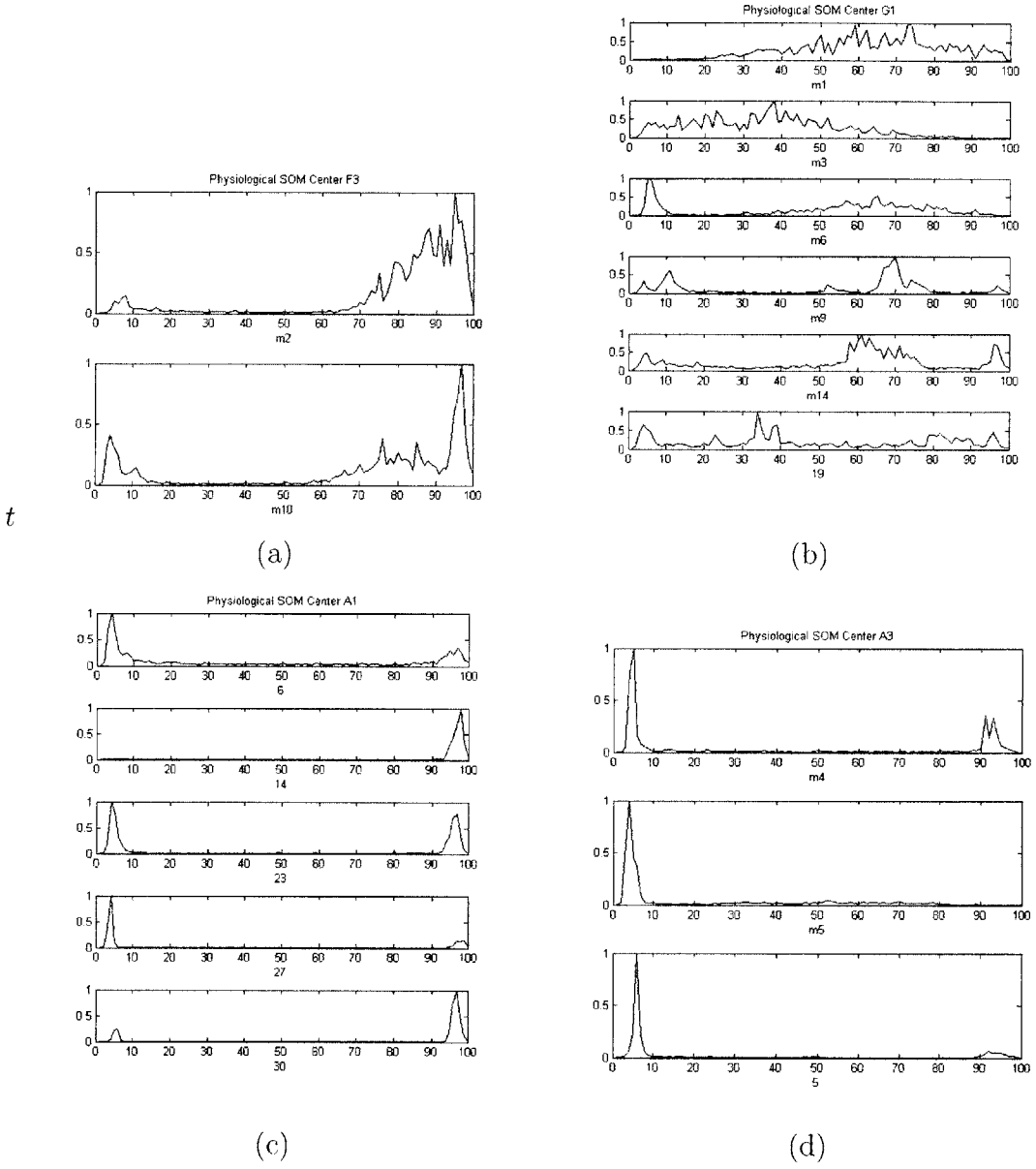


Figure 5-22: Band 3 morphologies for patients in center (a) "F3," (b) "G1," (c) "A1," and (d) "A3" of the SOM drawn from physiological feature space , shown in Fig. 5-15. As in the PCA-based SOM, we label centers by row and column; each row is assigned a letter (starting with A at the top) and each column is assigned a number (starting with 1 at the left).

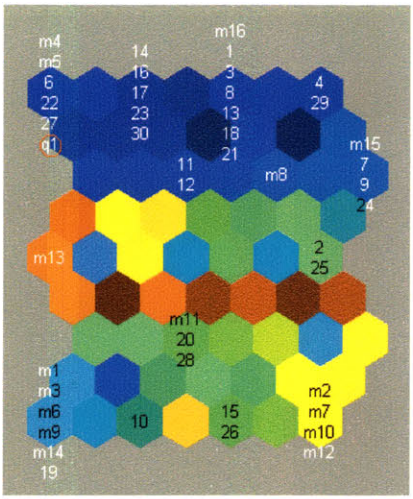


Figure 5-23: Projection of patient 5, as q1, onto physiological feature SOM trained without it. q1 appears in the upper-left portion of the SOM, circled in red.

Changes in SOM Appearance Across Training Sets

In both feature sets, removing individual patients for leave-one-out testing usually has little effect on SOM appearance. Often, the new map looks almost identicle to the one holding the excluded patient. The removal of a few patients lying in sparsely-populated red hexagons, however, cause variations in the coloring scheme within each region that alter the SOM's overall appearance. Nonetheless, in the worst cases in our data, we still identify "mostly MR" and "mostly no MR" regions as corresponding to the map incorporating the whole patient population. Classification remains tractable by relying more heavily on the arrangement of labels rather than the arrangement of colors.

Leave-one-out testing provides some intuition for the stability of our SOM representation to a changing data set. We might perform a more realistic assessment by removing half of our training data rather than one point. However, the small size of our patient population makes it likely that such a test would eliminate all the patients defining one or more important diagnostic regions. We gain some confidence in SOM stability based on the effects of removing individual outliers, which are most likely to effect SOM output and define the MR region in PCA-based feature space.

Maps from the physiological feature space vary predominantly in grid dimensions,

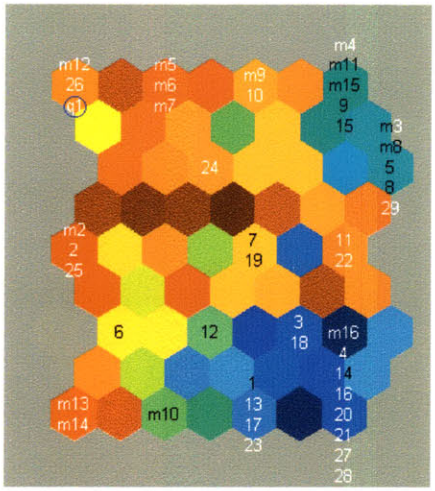


Figure 5-24: Projection of patient `m1`, as `q1`, onto PCA-based SOM trained without it. `q1` appears in the upper-left portion of the SOM, circled in blue.

when they change in appearance at all. We instuct the SOM algorithm to find a rectangular grid with roughly 20 centers. Depending on the training data, it outputs either a thin 3 by 7 grid, like that shown in Fig. 5-15, or a square-like 4 by 5 grid, like that shown in Fig. 5-23. Both forms retain the same general distribution of colors and labels. A red line divides all SOMs in the physiological feature space into two diagnostic regions. When we leave out patient `5` from the data set in Fig. 5-23, the resulting SOM continues to place the same patients in the most-populated MR center, at the bottom-left. The labels in the dark blue left edge of the full-population SOM are almost identical to those in the three dark blue hexagons in the top row of our new map.

Maps derived from the PCA-based feature space organize patient labels in a relatively consistent manner, although maps derived by removing some outlier patients vary considerably from Fig. 5-20 in their distributions of hexagonal colors. Neither Fig. 5-24 nor 5-25 contains the red edge confined to the upper-left portion of the SOM. However, the contents of this edge stay relatively close together in the new SOMs, as the labels are rotated and flipped along the grid. We can consider the top of each of the new SOMs to be a "mostly MR" region, following the rule established in Section 5.3.3 for the original PCA SOM. Fig. 5-24 seems to establish this MR region

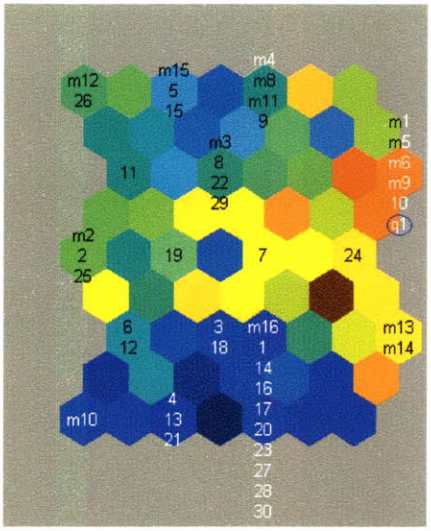


Figure 5-25: Projection of patient `m7`, as `q1`, onto PCA-based SOM trained without it. `q1` appears in the middle-right portion of the SOM, circled in blue.

with a dark red horizontal line separating lighter red hexagons above and below it. Patients without MR continue to cluster in a blue region at the bottom of each of our SOMs.

Analysis of the SOM in Fig. 5-20 helps explain the changing color distributions we observe for maps in PCA-based feature space, supporting our claims above about the harm of removing outliers. Across leave-one-out tests, the greatest changes in SOM appearance stem from removing MR files. Most patients with MR appear in red or light green in Fig. 5-20, indicating they are relatively distant from one another in the original feature space. Thus, the absence of one training point can make a significant difference in the optimal positioning of SOM centers in the feature space: centers can fall into more sparse or more dense areas than in the original SOM. New principal components also are likely to appear different when they no longer have to reflect the unusual morphology of a patient now removed from the training set. The new PCs can alter the patient distribution in the resulting feature space. Patients without MR often fall relatively close together in the feature space, as shown by the blue regions of Fig. 5-20. The removal of such patients from the training set causes little variation in SOM appearance, either in color or in label distribution, as seen in Fig. 5-26.

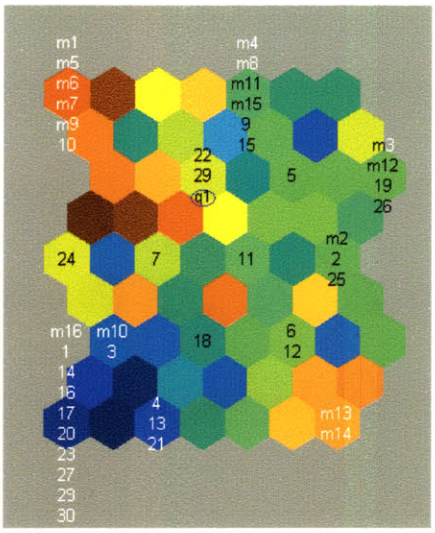


Figure 5-26: Projection of patient 8, as q1, onto PCA-based SOM trained without it. q1 appears in the upper-middle portion of the SOM, circled in blue.

"Leave-One-Out" Statistics

For each patient in our data set, we used our system to create SOMs in physiological and PCA-based feature space based on all patients except the selected one. We then projected the "left out" patient onto each SOM, seen as q1 on several example figures above. The author applied the SOM evaluation guidelines established in Section 5.1 to classify q1. If the patient fell in a "mostly MR" region, the classification was "MR;" if s/he appeared in a "mostly no MR" region, the classification was "no MR;" and if the region was unclear, the patient was classified as "MR." For each feature space, the author relied on the regions established in this chapter for the SOM built on the full data set, though the regions occassionally had to be rotated with the labels.

The author was able to classify correctly most patients with and without MR based on their projections onto SOMs. In our two feature spaces, human-driven classification correctly identifies 9 and 10 out of 16 MR patients respectively. The set of false negatives from the two spaces has little overlap. Thus, we can drive our false negative rate down to 2 or 1 out of 16 by choosing the most pessimistic classification, *i.e.*, if either SOM indicates the patient has MR, we classify the patient as "likely MR." Unfortunately, maximizing correct MR diagnoses inflates the false

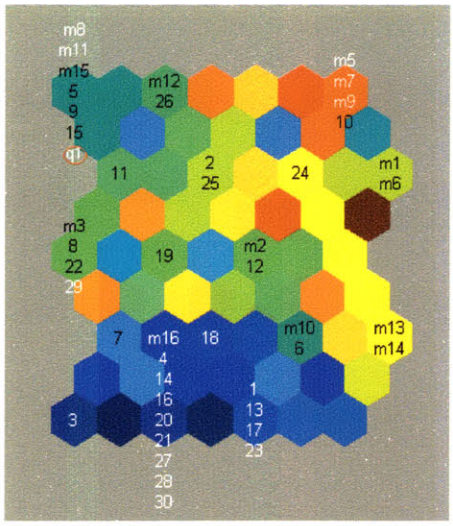


Figure 5-27: Projection of patient `m4`, as `q1`, onto PCA-based SOM trained without it. `q1` appears in the upper-left portion of the SOM, circled in red.

positive count to as many as 11 out of 30. We hope to drive false positives down in the future, but remain mindful that false positives cause much less harm than do the false negatives we already minimize.

Hard-to-Classify Patients

Each of our SOMs can fail to offer sufficient guidance for the classification of a particular patient. Thus, our system's performance in part depends on human judgement calls.

Two elements of human decision are involved in SOM-based diagnosis: the drawing of boundaries for diagnostic regions and the decision of how to handle outliers. In leave-one-out testing, we interpreted patient `m4`'s projection, as the label `q1`, into the upper-left corner of Fig. 5-27 to be unclear screening guidance. However, one can argue that the presence of 3 MRs in the same center—a large percentage of the MR population—should be sufficient to consider `q1`'s region as MR. Similarly, patient `19`'s position in the middle-right region of Fig. 5-28 may be considered a clear MR diagnosis, given proximity to `m3` and `m12`, a clear no MR diagnosis, given the shared hexagon with `25`, or uncertain, considering the difficulty of judging between the prior two possibilities based on coloring and location. Presuming patients `m4` and `19` are

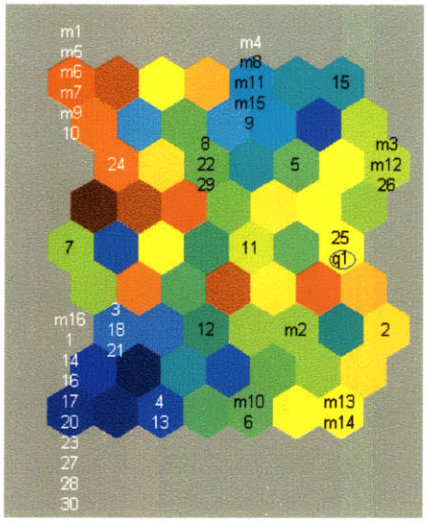


Figure 5-28: Projection of patient 10, as q1, onto PCA-based SOM trained without it. q1 appears in the middle-right portion of the SOM, circled in blue.

in uncertain territory during leave-one-out testing, we still must provide each a label. We favor the "pessimistic" choice, declaring MR when there is some doubt. This approach produces the correct answers for one patient, but misclassifies the other.

In a clinical setting, we would expect a physician to resolve the uncertainty based on further knowledge about the patient. The cardiologist enlisted to suggest classifications based on our acoustic recordings emphasized that screening for MR is more nuanced than binary classification through auscultation. Our presentation of a partial classification may make our system more suitable to the medical domain than are traditional binary classifiers, like thresholds and SVMs.

5.5 Physician-Based Classification

We recruited Dr. Collin Stultz, a cardiologist, to listen to a subset of the recordings used throughout this chapter to provide his assessment of each patient. We compare the classification provide by our system and by Dr. Stultz on these 13 patients with MR and 10 patients without MR.

Our system performs equivalent to or slightly worse than Dr. Stultz. Depending on the judgement of our human SOM-reader, SOM-based classification mislabels 5 or

6 patients, while Dr. Stultz mislabeled 4. Presuming the 5 error system, we have one fewer false negatives than he does. He misclassifies the one false negative in doubt in our system, patient `m4` discussed above. Our system also misclassifies Dr. Stultz's 2 false positive patients, in addition to making 2 further false positive assignments. On the whole, our two SOM visualizations generally provide the benefits, and the pitfalls, of a specialist's ear.

The abundance of comments Dr. Stultz provided, when originally requested only to indicate "has MR" and "send for echo" for each patient, emphasizes the complexity of medical decision making. Several times, Dr. Stultz commented that nobody should be sent for an echocardiogram on the basis of their heart sounds alone. We intend the degrees of diagnostic certainty permitted by SOMs to reflect realistically the *modus operandi* of medical science.

Comparison between the classification guidance from our system and from our cardiologist collaborator strongly suggests our system will perform well in a hospital screening. Our system makes only a few more errors than does the cardiologist, allowing most doctors to sharpen their understanding of a patient's heart sounds.

Chapter 6

Conclusions

Our assisted auscultation system provided significant help in identifying patients with mitral regurgitation (MR) based on their heart sounds. Analysis of the system blocks offered insights into the screening problem and suggested directions for further work.

6.1 Summary

Our system introduced and incorporated multiple tools that could be used in screening for MR. Physiological and PCA-based feature sets revealed distinct, clinically meaningful patterns in our data set. Self-organizing maps (SOMs) enabled us to visualize this structure, while suggesting a classification approach that may be useful to medical practice. In leave-one-out testing, the author used this system to classify patient with accuracy approaching that of a cardiologist.

This thesis introduced three novel feature sets: generic, physiological, and PCA-based. Each separated patients with and without MR from one another, although some sets were more effective than others. Physiological and PCA-based feature sets each allowed us to identify correctly different subsets of the MR patient population.

Physiological features drew the clearest boundary between our two classes of data. We distinguished between patients predominantly based on the presence or lack of significant mid-systolic energy. Further features allowed us to detect additional patients with MR. Incorporation of prior medical knowledge proved to be more useful

in interpreting prototypical systoles than was the application of purely statistical techniques.

Principle component analysis (PCA) also made important contributions to our screening task. Extracted principle components represented meaningful S2 and MR murmur shapes. Features based on these shapes organized patients into diagnostic regions in the SOM. However, MR patients spread out widely in the PCA-based feature space. As a result, removal of individual MR patients in leave-one-out testing changed the appearance of the MR region from one SOM to the next. Nonetheless, the separation between patients with and without MR remained and could be detected using the analysis techniques explained in Section 5.1.

Self-organizing maps provided visual intuition about the distribution of our patient population in each feature space, and the relation of new, undiagnosed patients to this population. SOM-based understanding permits a user to assess the reliability of our system for him/herself and provides guidance in screening for MR.

SOMs were useful in judging the utility of the physiological and PCA-based feature sets. For each set, we evaluated the ease of finding diagnostic regions on the map based on grid coloring and label placement, and sought features that maximized the number of patients captured in diagnostic regions. In another evluation technique, we examined the protoypical systoles of patients grouped together in the SOM. When the SOM places systolic shapes that are similar in appearance near each other, one has greater confidence in both the feature set and the clustering algorithm producing the map.

SOM-based screening performed better than other classification techniques we used on the same data set. In PCA space, the sparse distribution of data points in the "mostly MR" region led the radial basis kernel support vector machine (SVM) to overfit outlier "no MR" patients. In contrast, the SOM allowed the human operator to recognize the outlier patients as exceptions.

SOM-based classification produced different false negative classifications for the PCA-based and physiological feature sets. A doctor could attain a very small number of total false negatives with our system by assigning the MR designation to any patient

who fell in the "mostly MR" region of the SOM for at least one of our 2 feature spaces.

By combining classifications from physiological and PCA-based features, a human operator could use our system to classify patients with accuracy close to that of a cardiologist listening to the same files. Leave-one-out testing of our system produced as many false negatives as our collaborating cardiologist does. Our classifier also made 4 false positive errors, 2 more than those made by the cardiologist. Because we intended to aid primary care physicians, who have poorer auscultation skills than a specialist, we were encouraged by the results of our comparitive study.

6.2 Future Work

Much work remains to optimize the assisted auscultation system and to improve its utility. We must recruit medical practitioners to test our SOM visualizations for intelligibility and to test the classification performance of the SOM-doctor combination. We should explore improvements in systole visualization and feature extraction blocks. We also should search for optimal SOM formation parameters for our data set. Beyond our current data, we may test the system over broader populations and explore screening for conditions besides MR.

A user test is the most important step to test the benefits of SOMs for doctors lacking a technical background. Several medical practitioners, including cardiologists and primary care physicians, should be enlisted to suggest classifications for our patients based on acoustic recording, SOM displays, or both.

We should explore a wide range of settings for the number of SOM centers, to observe its effect on classification performance and on intelligibility. Preliminary testing indicates a decrease in centers results in improvements in all aspects of the system, but this remains to be thoroughly investigated.

Chapter 5 suggests our other system blocks also may benefit from tuning. For example, in the systole visualization block, we should seek to reduce the variability in the formation of prototypical systole, revisiting the non-deterministic selection of individual systoles to be incorporated into the prototypical.

In addition to improving our performance on our current 46-patient data set, we should explore how well our system scales to larger populations, handles recordings from multiple positions on the body, or distinguishes between multiple heart conditions. As we increase the size of our training set, we will have to modify the SOM displays to keep them intelligible. Our current system lists all patients individually, causing labels in some SOMs, such as Fig. 5-20, to flow past the bounds of the map and some to appear misleadingly over hexagons with which they are not associated. Increasing the number of centers may avoid overflow, while revealing finer structures visible from a larger data set. Alternatively, for each center, we may list one `MR` and one `no MR` label and pick the size of each label according to the number of files of the associated class falling into the center.

Bibliography

[1] Dr. Robert Levine, Private Communication.

[2] Dr. Collin Stultz, Private Communication.

[3] Exercise physiology digital image archive: blood flow through the heart. http://www.abacon.com/dia/exphys/one.html. [Online document] [2006 May 25].

[4] Principal components analysis - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Principal_components_analysis. [Online document] 19 May 2006, [2006 May 22].

[5] Som toolbox 2.0 documentation. http://www.cis.hut.fi/projects/somtoolbox/documentation/. [Online document] 18 Mar 2005, [2006 Feb 13].

[6] M. Akay. Wavelet applications in medicine. *IEEE Spectrum*, 34(5):50–56, 1997.

[7] M. Akay, J.L. Semmlwo, W. Welkowitz, M.D. Bauer, and J.B. Kostis. Detection of coronary occlusions using autoregressive modeling of diastolic heart sounds. *IEEE Transactions on Biomedical Engineering*, 37(4), 1990.

[8] E. Braunwald, editor. *Heart disease: a textbook of cardiovascular medicine*, chapter 2. W.B. Saunders Company, 2001.

[9] L.-G. Durand and P. Pibarot. Digital signal processing of the phonocardiogram: review of the most recent advancements. *Critical Reviews in Biomedical Engineering*, 23(3-4):163–219, 1995.

[10] E. Gabrilovich and S. Markovitch. Text categorization with many redundant features: Using aggressive feature selection to make svms competitive with c4.5. In *Proceedings of The Twenty-First International Conference on Machine Learning*, pages 321–328, Banff, Alberta, Canada, 2004. Morgan Kaufmann.

[11] L. Goldman and D. Ausiello, editors. *Cecil Textbook of medicine*, chapter 48. W.B. Saunders Company, 22nd edition, 2003.

[12] T. Hastie, R. Tibshirani, and J. Friedman. *Unsupervised learning*, chapter 14, pages 437–508. Springer, 2001.

[13] M. Jung. Automated auscultation: using acoustic features to diagnose mitral valve prolapse. Master's thesis, Massachusetts Institute of Technology, 2004.

[14] T. Kohonen. *The basic SOM*, chapter 3. In [15], third edition, 2000.

[15] T. Kohonen. *Self-organizing maps*. Springer, third edition, 2000.

[16] T.S. Leung, P.R. White, J. Cook, W.B. Collins, E. Brown, and A.P. Salmon. Analysis of the second heart sound for diagnosis of paediatric heart disease. In *Proc. IEE Science, Measurement and Technology*, pages 285–290, 1988.

[17] Z. Syed. Mit automated auscultation system. Master's thesis, Massachusetts Institute of Technology, 2003.

[18] Z. Syed, D. Leeds, D. Curtis, F. Nesta, R.A. Levine, and J. Guttag. A framework for the analysis of acoustical cardiac signals. *IEEE Transactions on Biomedical Engineering*. Submitted for Publication in October 2005.

[19] J.M. Vukanovic-Criley, S Criley, C.M. Warde, J.R. Boker, L. Guevara-Matheus, W.H. Churchill, W.P. Nelson, and J.M. Criley. Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study. *Archives of internal medicine*, 166(6):610–6, 2006.

[20] X. Zhang, L. Durand, L. Senhadj, and J.-L. Coatrieux H.C. Lee. Time-frequency scaling transformation of the phonocardiogram based on the matching pursuit method. *IEEE Transactions on Biomedical Engineering*, 45(8):972–979, 1998.