

A Framework for the Analysis of Acoustical Cardiac Signals

Zeeshan Syed, Daniel Leeds, Dorothy Curtis, Francesca Nesta, Robert A. Levine, and John Guttag

Abstract—Skilled cardiologists perform cardiac auscultation, acquiring and interpreting heart sounds, by implicitly carrying out a sequence of steps. These include discarding clinically irrelevant beats, selectively tuning in to particular frequencies and aggregating information across time to make a diagnosis.

In this paper, we formalize a series of analytical stages for processing heart sounds, propose algorithms to enable computers to approximate these steps, and investigate the effectiveness of each step in extracting relevant information from actual patient data. Through such reasoning, we provide insight into the relative difficulty of the various tasks involved in the accurate interpretation of heart sounds. We also evaluate the contribution of each analytical stage in the overall assessment of patients.

We expect our framework and associated software to be useful to educators wanting to teach cardiac auscultation, and to primary care physicians, who can benefit from presentation tools for computer-assisted diagnosis of cardiac disorders. Researchers may also employ the comprehensive processing provided by our framework to develop more powerful, fully-automated auscultation applications.

Index Terms—Cardiac screening, heart murmurs, auscultation, automated diagnosis

I. INTRODUCTION

IN the United States over 23 million adults (approximately 11.2% of the total adult population) have diagnosed cardiac disorders. In 2002, more than 4.4 million people in the U.S. were hospitalized owing to cardiac conditions [1].

Manuscript received September 17, 2005. This work was supported in part by the Center for Integration of Medicine and Innovative Technology (CIMIT), the Deshpande Center for Technological Innovation, the Aetna Foundation, and the MIT Project Oxygen Partnership.

Zeeshan Syed, Dorothy Curtis and John Guttag are with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (phone: 617-253-1996; fax: 617-253-8460; e-mail: zhs@csail.mit.edu).

Daniel Leeds was with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with Carnegie Mellon University, Pittsburgh, PA 15213 USA.

Francesca Nesta., was with the Massachusetts General Hospital, Boston MA 02114 USA. She is now with the Boston Medical Center, Boston, MA 02118 USA.

Robert A. Levine is with the Massachusetts General Hospital, Boston, MA 02114 USA, and Harvard Medical School, Boston, MA 02115 USA.

Copyright (c) 2006 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

The first line of defense against cardiac disease is the annual physical exam. The primary care physician uses an ECG to assess the electrical activity of the heart, and a stethoscope to listen for other kinds of problems, e.g., malfunctioning valves. The process of interpreting the sounds produced by the heart is called cardiac auscultation. Different mechanisms of heart disease are known to produce distinct blood flow disturbances varying in velocity, Reynolds number and timing [2]. These distinguishing aspects are expressed in the corresponding acoustic findings of frequency content, loudness and timing; and provide the rationale for auscultatory analysis. Unfortunately, detecting relevant symptoms and forming a diagnosis based on sounds heard through a stethoscope is a difficult skill that takes years to acquire and refine. Most physicians never master it for a variety of reasons: the heart sounds of interest are often of short duration, and may be separated from one another by less than 30 ms [5], the sounds indicating cardiac disorders are typically much quieter than other heart sounds, and there is considerable beat-to-beat variation.

Quantifying the impact of the difficulty of cardiac auscultation is hard. There is no way of determining how many problems are missed. What is known is that it leads to many unnecessary referrals. It has been estimated that 80% of patients referred to cardiologists as the result of auscultatory exams have benign heart murmurs or normal hearts [5,6]. These cases represent a considerable cost to the medical system since a visit to a cardiologist (including associated echocardiography) runs anywhere from \$300 to \$1000 in the United States [5]. In addition to the monetary costs, these false positives constitute a significant waste of time for both patients and cardiologists, and are also the source of much unnecessary emotional anxiety for patients and their families.

Skilled cardiologists interpret heart sounds by carrying out a sequence of steps, some of which may be performed at a subconscious level. These include discarding beats that are significantly corrupted by noise or not deemed to have sufficient diagnostic content, selectively tuning their hearing to particular frequencies corresponding to diseased activity, and aggregating information across the examination to make a diagnosis.

In this paper, we present an analytical perspective on cardiac auscultation. We focus on disorders such as valvular disease and stenosis that are typically detected using auscultation rather than on disorders that are typically detected

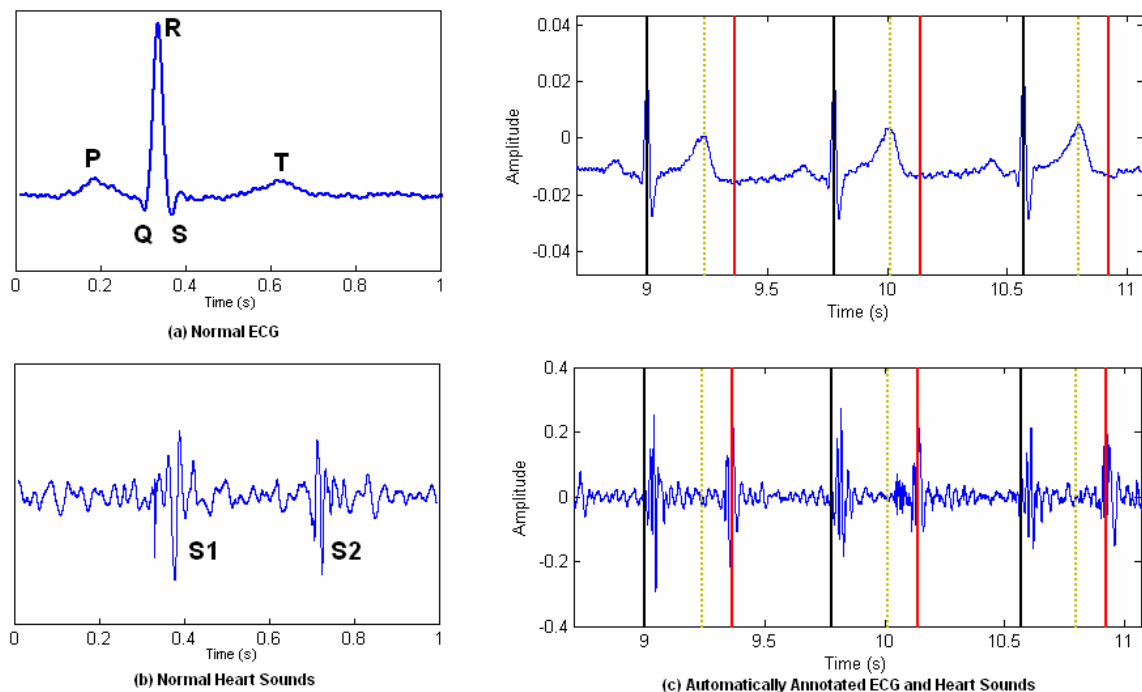


Fig. 1. Characteristic ECG and acoustical cardiac activity is shown in (a) and (b) respectively for time-correlated signals. In (c), the techniques described in Section 3 are used to automatically determine locations of each R wave (solid black), T wave (dotted brown) and S2 (solid red) in a segment of simultaneously recorded ECG and audio.

using ECG. An implicit assumption of our work, which is used in segmenting the signal, is that subjects have hearts that are normal from an electrophysiological perspective. We formalize a series of analytical stages that are used to process audio signals from the heart, and propose algorithms to enable computers to approximate these steps in a fully automated manner. Through such a line of reasoning, we provide an appreciation of the relative difficulty of the various tasks involved in the accurate interpretation of heart sounds. We also describe software that implements each stage, and experimentally evaluate the potential contribution of each analytical stage in the overall assessment of patients.

Background on ECG and acoustical cardiac signals, as well as previous work in the area of automated and semi-automated analysis of heart sounds, is presented in Section 2. This is followed by a description of the system architecture and the individual components of our framework in Section 3, a review of our testing methodology and the corresponding results in Sections 4 and 5, and conclusions in Section 6.

II. BACKGROUND

A. ECG and Heart Sounds

The electrocardiogram (ECG) [4] provides a non-invasive measurement of the electrical activity of the heart. A typical ECG tracing, corresponding to a single cardiac cycle is presented in Figure 1a. Distinct electrophysiological events appear as disturbances in the ECG signal. The P wave in Figure 1a corresponds to the electrical excitation

(depolarization) of the top two atrial chambers of the heart. The P wave is associated with blood being pushed by atrial contraction into the lower two ventricular chambers. The Q, R and S waves together form the QRS complex, which is associated with contraction of the ventricles due to ventricular depolarization. This results in blood being pushed out of the heart into arterial vessels. The T wave corresponds to repolarization of the ventricles, which restores the heart tissue to the normal state and allows the ventricles to relax prior to the next cardiac cycle. (Atrial repolarization is typically concealed by the higher-amplitude QRS complex in the normal ECG.)

The electrical activity of the heart produces mechanical effects that manifest themselves as acoustical signals [4]. Figure 1b shows time-correlated audio with the ECG in Figure 1a for a normal heart. The first heart sound, called S1, occurs shortly following the R wave. It is produced as a result of ventricular contraction causing blood to flow back towards the atria, shutting the AV valves between the chambers. The second heart sound, S2, can be heard at the end of the T wave. This is produced by the relaxation of the ventricles causing blood to flow back into these chambers from the arteries, shutting the valves between the ventricles and the arterial vessels. In each case, the closing of valves is associated with vibrations that produce sounds.

In the acoustic signal, the period from S1 to S2 is known as systole, while the S2-S1 phase corresponds to diastole. In ECG, the definition of systole varies slightly, with the QRS complex corresponding to ventricular contraction being used as the start of systole. In the remainder of this paper, we

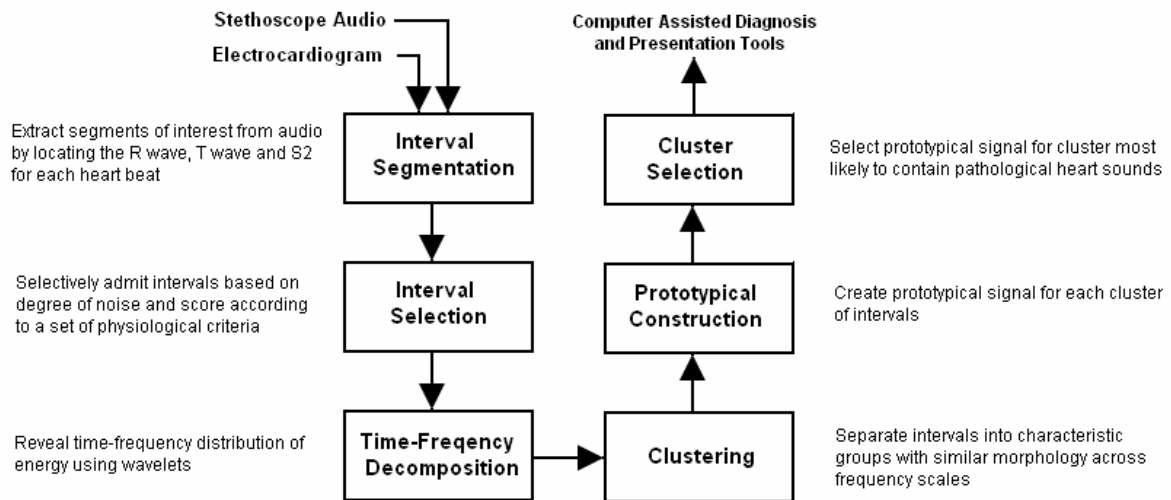


Fig. 2. Modular representation of auscultation adopted by the MAAS Framework

choose to define systole as the period between the QRS and S2, with the remainder being labeled as diastole. This choice is motivated by the impulse like characteristic of the QRS and the relatively noise-free nature of the ECG signal in general. These factors allow for a more precise determination of the start of systole than can be located by searching for S1 in the audio signal. Figure 1 illustrates this effect.

B. Related Work

Researchers have used various approaches to analyze acoustic cardiac signals, and there is an extensive literature on software to classify heart sounds [8]-[15]. Most of these efforts focus on segmenting the audio signal into individual beats, identifying the presence of noise, and analyzing the spectral content of the resulting beats.

The work described in [9] and [12] depends on manual segmentation with experts annotating the raw audio signal with cardiac timing information. Various approaches have been proposed to automate this process. In [13], information in the ECG is used to detect the R wave, with the first heart sound declared to last for 110 ms from the QRS complex and the second heart sound detected by searching for a peak in the audio signal following this period. [14] describes an approach to detect S1 and S2 from the acoustical signal alone, making use of the assumption that the S1-S2 separation is shorter than the S2-S1 distance. In [11], the time-frequency decomposition of the acoustical signal is used instead of the raw audio. Heart sounds are decomposed using wavelets and half second windows are passed to a single layer perceptron trained to detect cardiac artifacts.

The problem of noise is explicitly addressed in [8] and [11]. In [8], the noise detection process starts by dividing systole into two segments, containing the first and second heart sound respectively. In each segment the variance is computed. All systoles with variance in either segment significantly above a threshold value are discarded as being noisy. The threshold value used depends on the smallest variance calculated across

all systoles. The method described in [11] detects noise by passing the wavelet coefficients of half second windows of audio to single layer perceptrons that have been previously trained on clean and noisy data. [11] also presents an interesting parallel architecture, in which segmentation and noise detection proceed simultaneously. This differs from the more usual approach of organizing processing blocks sequentially.

The most popular technique for spectral analysis of the beats resulting from segmentation is wavelet-based time-frequency decomposition ([8], [9], [11]). An alternative, proposed in [12] and [13] is to identify the frequency content of a signal using an FFT with gross timing information provided by the segmentation of the signal.

In the remainder of this paper, we build upon this existing work, suggesting an alternate approach to segmentation, and introducing processing blocks beyond segmentation, noise-detection and spectral analysis, i.e., clustering, prototypical construction and use of physiological criteria for interval selection.

III. SYSTEM DESCRIPTION

A. Overview

Figure 2 presents the architecture of the MIT Automated Auscultation System (MAAS) Framework. Each functional block in the diagram is associated with a distinct analytical stage of auscultation.

The first processing stage is *interval segmentation*. Signals from an electronic stethoscope are annotated with activity from a simultaneously recorded ECG stream. This facilitates identification and extraction of audio sections of interest for each cardiac cycle. Since the diagnostic content of intervals associated with some heart beats may be greater because of physiological variations that influence the audibility of heart sounds across beats, output from the interval segmentation

stage is passed through a filter that performs *interval selection* and selectively admits each interval based on its score according to a set of physiologically based criteria. This stage also rejects noisy intervals. Next, energy distribution characteristics of these admitted intervals are examined using *time-frequency decomposition* techniques. Groups of intervals with similar morphology across different frequency scales are then isolated into *clusters*. For each statistically significant cluster, the intervals are pooled together to create representative *prototypical intervals*, which remove artifacts due to noise while emphasizing recurring pathological signals. This allows for a robust, compact representation of the acoustical cardiac activity of the patient. The prototypical candidate most likely corresponding to a diseased cluster of intervals is identified automatically via a *cluster selection* mechanism, and this signal is used for disease assessment.

B. Interval Segmentation

Segmentation of raw audio into intervals corresponding to different stages of the cardiac cycle is essential to correlate heart sounds to underlying electrophysiological activity because pathological artifacts across distinct cardiac ailments may possess similar acoustical properties. In these cases, successful discrimination of a disorder relies heavily on timing information and knowledge of the specific intervals during which abnormal activity can be discerned. Systolic and diastolic murmurs are often distinguished in this manner, by determining the segments during which they are consistently heard.

Initially, information in the ECG signal is used to locate the R wave and the T wave for each heart beat. S2 is then identified by using information in the audio signal. Systole, as defined in Section 2, corresponds to the regions of the signal between the R wave-S2 pair for each heart beat, while the rest of the recording is diastolic.

To find R waves we apply a modified version of the algorithm in [19] to the ECG signal. This technique is robust in the face of electromyographic noise and powerline interference [20]. Given an ECG signal $X_e[n]$ for $1 \leq n \leq N$, an amplitude threshold θ is determined such that:

$$\theta = 0.4 \max_n(|X_e[n]|) \quad (1)$$

The data is then rectified and clipped to synthesize $X_e'[n]$ for $1 \leq n \leq N$:

$$X_e'[n] = \begin{cases} |X_e[n]| & |X_e[n]| > \theta \\ \theta & |X_e[n]| \leq \theta \end{cases} \quad (2)$$

An R wave is then declared at each position where:

$$X_e'[n+1] - X_e'[n-1] \geq \frac{1}{3} \quad (3)$$

T waves are located by examining the ECG signal between

consecutive R waves. Denoting the R and T waves for beat i as $R[i]$ and $T[i]$ respectively, and defining $RR[i] = R[i+1] - R[i]$, we declare a T wave at positions in the ECG signal such that:

$$T[i] = \arg \max_n (|X_e[n]|) \quad (4)$$

$$R[i] + \frac{RR[i]}{4} \leq n \leq R[i+1] - \frac{RR[i]}{2}$$

In the case of S2, the ECG does not provide any clear indication of the end of systole. Instead, the second heart sound serves as the means for determining the endpoint of systolic activity. Locating S2 is challenging because:

- Severe systolic murmurs, particularly those that take place late during late systole, often produce energy patterns similar to the second heart sound,
- S2 is often composed of multiple distinct phases (i.e., "split S2"),
- Artifacts such as clipping (e.g., because of movements of the stethoscope during the recording) can be confused with S2, and
- There is significant variation in systolic duration across patients

To address these issues, we develop a mechanism to estimate systolic length on a per-patient basis. We start off by using knowledge that the T wave always precedes S2 to guide our search. The audio signal is examined over a window of 150 ms following the T wave for a peak in amplitude.

Denoting the S2 candidate for beat i by $S2_c$ and the audio signal by $X_a[n]$ for $1 \leq n \leq N$, we obtain:

$$S2_c[i] = \arg \max_n |X_a[n]| \quad (5)$$

$$n = T[i], \dots, T[i] + 150ms$$

Some of the candidate second heart sounds isolated in this manner correspond to late systolic murmurs and peaks due to artifacts. To deal with this, we make use of the fact that although the length of systole may vary significantly across patients, it is relatively uniform for most patients over the course of a single recording. On the other hand, the location of artifacts varies randomly. Lowpass filtering the sequence of data points representing the distance between each R wave and the subsequent S2 candidate, provides a useful patient-specific estimate of the duration of systole (denoted by $RS2len$):

$$RS2len = filter \left(\begin{bmatrix} S2_c[1] \\ S2_c[2] \\ \vdots \end{bmatrix} - \begin{bmatrix} R[1] \\ R[2] \\ \vdots \end{bmatrix} \right) \quad (6)$$

Using this value, we search for a peak in audio amplitude corresponding to S2 for beat i in the vicinity (± 10 ms) of $R[i] + RS2len$, i.e.:

$$S2[i] = \arg \max_n |X_a[n]| \quad (7)$$

$$n = R[i] + RS2len - 10ms, \dots, R[i] + RS2len + 10ms$$

In the case of split S2, the peak with greater amplitude is selected. Figure 1c presents the results of our approach, by showing time-correlated ECG and audio signals along with the R waves, T waves and S2s automatically detected.

C. Interval Selection

The interval selection stage determines which of the segmented intervals should be analyzed further. This decision is based on two factors: degree of corruption by noise, and score using physiologically-based criteria aimed at assessing the diagnostic value of each interval

Given the enormous set of cardiac disease signatures, it is difficult to discriminate reliably between noise and energy caused by abnormal heart activity. For this reason, the interval selection stage focuses only on discarding intervals that are significantly corrupted by noise (i.e., with high energy artifacts that are considerably louder than both S1 and S2, and can readily be attributed to noise). Intervals with lower levels of noise are retained in the expectation that they may still contain valuable diagnostic information. The algorithms described in later sections make use of various aggregation techniques to operate robustly on these signals.

Intervals are also admitted based on their scores according to a set of physiologically-based criteria. For example, systolic segments preceded by long diastoles are given greater weight than those following short diastolic segments. The rationale behind this criterion is that a longer diastole allows the ventricles more time to fill with blood. This leads to an increase in the volume of blood passing through the heart during systole, which in turn produces more audible cardiac sounds.

In essence, this step is geared towards selecting the intervals that are most likely to display the presence of heart disease of the kind we are trying to detect, and therefore warrant greater attention.

D. Time-Frequency Decomposition

Cardiac disorders related to valvular disease and stenosis are typically diagnosed by examining the timing, amplitude and spectral character of heart sounds. Timing information allows for events to be correlated to the underlying activity of the heart. Loudness variations correspond to an increase or decrease in the amplitude of the signal. Spectral irregularities reveal changes in pitch associated with shifts of energy to higher or lower frequencies due to abnormalities such as the turbulent flow of blood during regurgitation.

The emphasis on evaluating signals in terms of time, amplitude and frequency suggests the analysis of heart sounds through time-frequency decomposition. We adopt a wavelets-based approach for this purpose, which employs short windows to isolate signal discontinuities and long windows to obtain detailed frequency analysis of gross features. As

opposed to filter banks, which exhibit fixed resolution at all locations in the time-frequency plane, wavelet multiresolution analysis allows for better time-resolution at higher frequencies, and better frequency-resolution at low frequencies. This allows for more fine-grained detection of the short bursts of high-frequency energy frequently associated with heart murmurs. In addition, wavelets also provide a sharp separation of the baseline S1-S2 signal from pathological activity. This is essential since the baseline S1-S2 signal is typically two to three orders of magnitude louder than abnormal high-frequency sounds (Figure 3). Improved frequency resolution at low frequencies prevents energy in abnormal ranges from being colored by energy in the baseline S1-S2 signal, which would otherwise conceal important spectral trends and reduce visibility of diagnostic signatures.

We decompose each segmented interval $X_i[n]$ for $1 \leq n \leq N_i$ using the wavelet transform:

$$C_i(s, p) = \sum_{n=1}^{N_i} X_i[n] \psi \left[\frac{n-p}{s} \right] \quad (8)$$

The mother wavelet ψ in (8) corresponds to a Daubechies scaling function [21] of order 5. For computational efficiency, we use the implementation described in [16] with dyadic scales and positions to achieve a level 9 decomposition of $X_i[n]$. A 2-channel dyadic filter splits the spectrum into half at each iteration. The mother wavelet ψ at scale s then obeys the form $2^{-s/2} \psi[2^{-s}n-1]$, with the output being reconstructed at every level to ensure that the activity isolated in each band is of equal length. The decomposition of $X_i[n]$ at levels 6 to 9 corresponds to the approximate frequency ranges 75-150 Hz, 150-300 Hz, 300-600 Hz and 600-1200 Hz. To avoid destructive interference in the prototypical interval calculation stage discussed later in this section, we output the absolute magnitude of all signals at this stage.

E. Clustering

Depending on the severity and nature of a heart disorder, clicks and murmurs may not be audible on every beat. It is reasonable to expect that some beats in a recording will sound normal whereas others will contain sounds generated by irregular heart function. We therefore attempt to group intervals together based on similarities between their time-frequency characteristics. This clustering is geared towards isolating intervals that sound different from others.

Systolic and diastolic segments are clustered using a K-means clustering algorithm [22] with $K=2$ and features chosen to be the positions relative to the R wave of peaks in each frequency band. The algorithm separates the intervals into two clusters, one of which is likely to correspond to normal sounding segments and the other to those that might contain pathological heart sounds. Noisy beats can be recognized as outliers and discarded [7].

F. Prototypical Construction

In order to observe the characteristic trends persisting

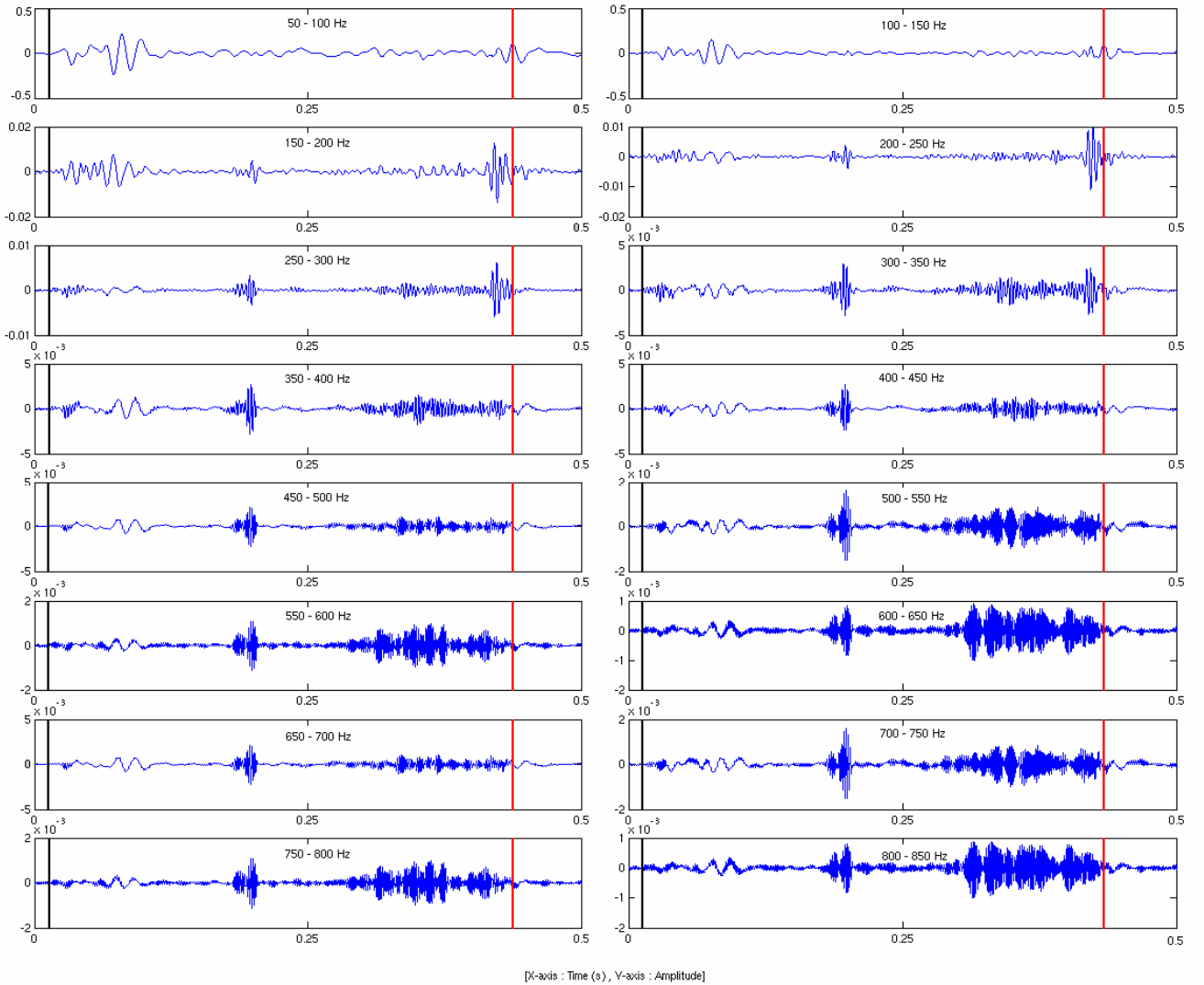


Fig. 3. Evolution of energy across frequencies for an exemplary patient suffering from mitral regurgitation. The presence of increased systolic energy due to regurgitant flow of blood between S1 (left black line) and S2 (right red line) can be seen most clearly at higher frequencies. Differences in the intensities of low frequency and high frequency sounds are reflected by the scales for the 50-100 Hz plot (-0.5 to 0.5) at the top left and the 800-850 Hz plot (-0.001 to 0.001) at the bottom right.

among the majority of the intervals in any cluster, we develop mechanisms to aggregate information across multiple segments [23]. A representative biological signal is created for each cluster, merging content across both time and amplitude axes. Events within this hypothetical prototype occur at averaged locations from a fiducial point of alignment, e.g., the R wave, and are smoothed out in amplitude based on recurring observations. One of the key advantages of pooling multiple intervals in this manner is that it allows us to derive a representation of the underlying cardiac activity while discarding noise.

Prototypical construction proceeds by hierarchically collapsing pairs of observations into single intervals. The process of merging two signals, e.g., $X_1[n]$ and $X_2[n]$ for $1 \leq n \leq N$, proceeds by first aligning the sequences through dynamic time-warping (DTW) [25]. Samples along contiguous edges of the alignment path K are then compacted by using amplitude information from both signals. Without loss of generality,

assuming that a sample from $X_1[n]$ at time n_1 is aligned with τ consecutive samples from $X_2[n]$ beginning at time n_2 , the $\tau + 1$ samples across both signals are merged through:

$$\frac{X_1[n_1]}{2} + \frac{1}{4} [X_2[n_2 + 2i - 1] + X_2[n_2 + 2i]] \quad i = 1, \dots, \varphi \quad (9)$$

Where φ is set to be the ceiling or floor of $\tau/2$ depending on whether τ is even or odd. A property of (9), proved in [23], is that the resulting signal produced by this pair-wise merging of content always has the same length as the original signals, allowing hierarchical combination of intervals. Intuitively, this occurs because (9) halves the number of samples. A discussion of optimizations to improve the quality of the resulting signal, which we employ but do not discuss here, is also deferred to [23].

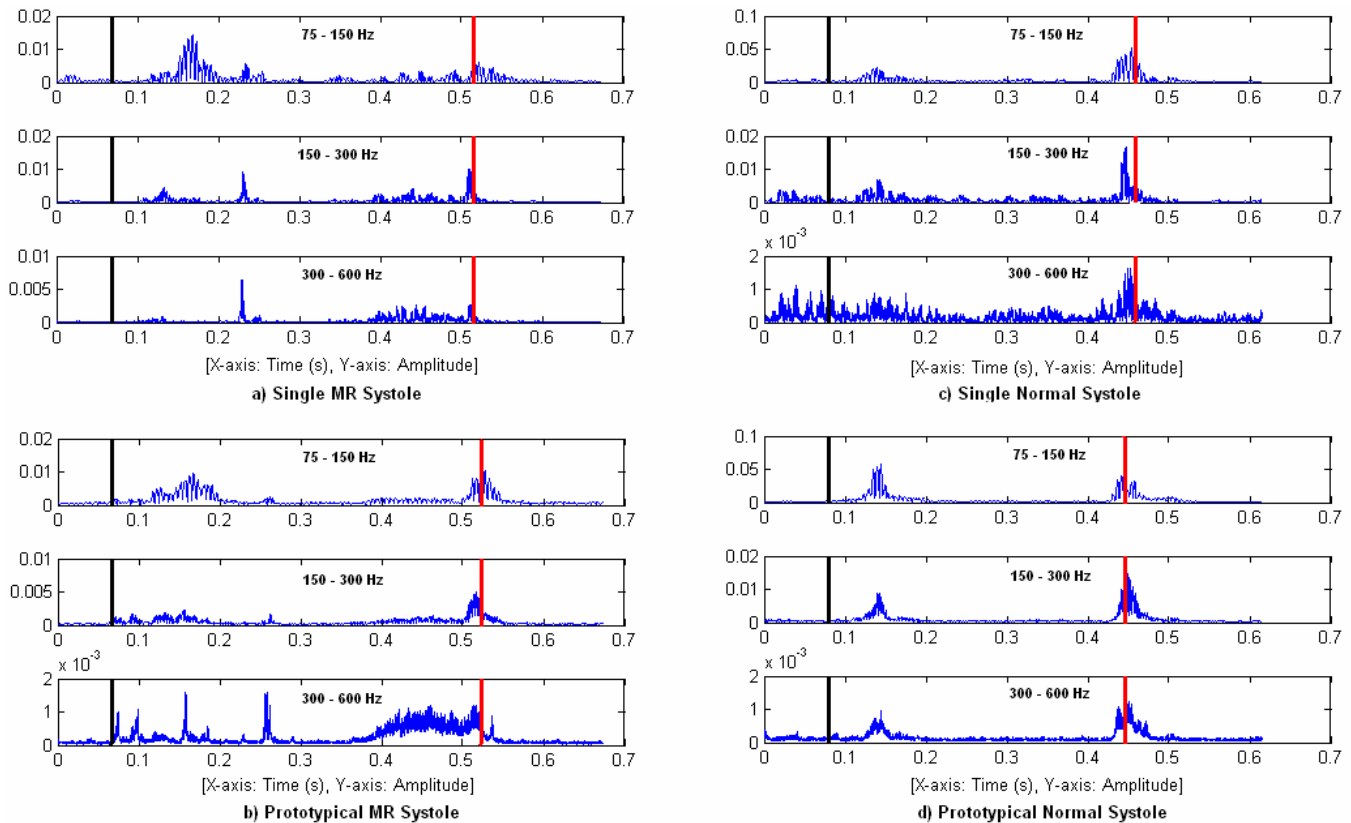


Fig. 4. Prototypical systoles are compared to single beats for patients with MR and with normal hearts. In (b), the prototypical systole shows the presence of MR more clearly, than in (a). The prototypical construction process preserves this information, as well as sharp clicks that recur across several beats close to S1. For the normal heart, the prototypical systole in (d) shows the absence of any clicks or high frequency murmurs. This is because the activity seen in single beats does not persist across beats. The black line (left) in all plots indicates the R wave, while the red line (right) corresponds to S2.

Figure 4 provides an example of the use of prototypes to perform more robust analysis of heart sounds. For the patient with MR, the prototype reveals abnormal activity more visibly, and in particular, makes it easier to distinguish the particular last-systolic characteristics of the murmur. Conversely, for the normal heart, the prototype removes noise and emphasizes the absence of pathological activity.

G. Cluster Selection

The cluster selection stage reduces the number of clusters to one. As a first step, if a cluster does not contain a statistically significant number of intervals (this value is currently set to eight) it is discarded. Otherwise, if both clusters have eight or more intervals, the cluster for which the prototypical interval has greater energy content in frequency bands other than those corresponding to the S1-S2 baseline signal is chosen.

The decision to bias the cluster selection process in favor of clusters with greater energy at higher frequencies is based on the rationale that heart disorders are typically associated with clicks and murmurs, both of which are expected to contribute additional abnormal energy content in those bands.

IV. TESTING METHODOLOGY

This section provides a quantitative evaluation of the impact of each processing stage discussed in Section 3. Assessing the

significance of each module in terms of its effect on the classification of cardiac signals is complicated by two distinct factors.

Firstly, measuring the effect of processing on classification is not a well-defined problem. There are a large number of cardiac disorders, a myriad of possible classification schemes, and an unbounded number of possible feature sets. Spanning this universe of possibilities is an intractable problem. Our initial work in this area therefore examined the effect of processing on the classification accuracy of a sample system to detect one problem, mitral regurgitation (MR). We evaluated the significance of the different processing modules within the context of this disorder and developed a case for the results of our study generalizing to a large set of other diseases. We also fixed the classification technique and feature set, as discussed below.

A secondary complication is that the processing blocks are not always independent and cannot always be evaluated in isolation. Some analytical stages have a strong dependence on other modules:

- The creation of prototypical signals requires the time-frequency decomposition of intervals, since aggregation of information occurs in both time and frequency [23],
- The clustering and cluster selection stages only make sense when both are used, and

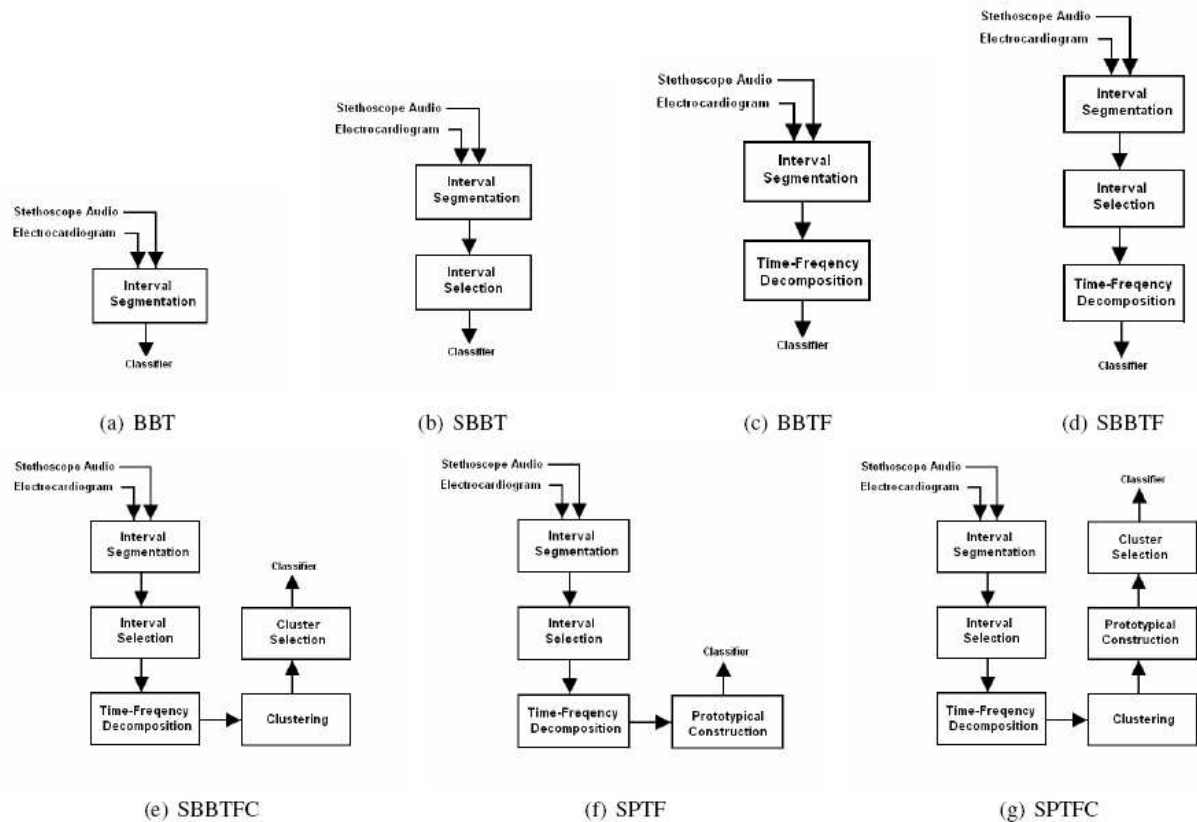


Fig. 5. Evaluated configurations of the MAAS Framework processing blocks (S = selected beats, BB = beat by beat, P = prototypical beat, T = time domain analysis, TF = time-frequency domain analysis, C = clustering)

- The clustering stage requires time-frequency information to group intervals with similar morphology across frequency scales.

We evaluated independent modules by determining their incremental contribution to classification accuracy. The effect of dependent processing blocks chained together was gauged by comparing the cumulative effect to the sum of its independent parts.

The dataset for our study consisted of 39 de-identified adult test subjects. 11 suffered from MR (roughly an equal number of mild, moderate and severe cases were included), 15 were found to have benign murmurs and the remaining 13 represented control subjects with normal hearts. Heart sounds were recorded simultaneously with ECG using a Meditron electronic stethoscope, which connected directly to a PC and sampled signals at 44,100 Hz with 16 bit quantization. Anti-aliasing filtering and down sampling to 4,096 Hz was performed prior to any processing block, to compromise the computational demands. A total of 30 to 40 second recordings were made at several pre-determined auscultation sites. In all but two cases, recordings from the left apex with patients in a supine position were chosen. In the two cases where there was significant corruption of the original signal due to noise, recordings from the left parasternal space were used instead. All diagnoses were verified through echocardiographic examination at the Massachusetts General Hospital.

To evaluate the contribution of each component of the

MAAS Framework, classification accuracy was calculated while experimenting with different combinations of the processing blocks. The configurations shown in Figures 5(a)–5(g) were examined.

In the experiments, a simple classifier was used to detect MR. It was based on discriminating the presence of increased energy during systole, possibly caused by the turbulent flow of regurgitant blood during that period. The features chosen for this purpose were:

- The position of the maximum amplitude peak relative to S2, and
- Energy during systole normalized by the amplitude of S2

For time domain classification, both features could be computed in a straightforward manner based on the amplitude of the segmented audio intervals. In the case of time-frequency domain classification, only activity in higher frequency bands was considered, and the maximum values across these bands were chosen as final feature values.

Although we calculated both features, for our dataset, we found that only the position of the maximum amplitude peak relative to S2 was useful for the purpose of distinguishing MR from benign murmurs and normal heart sounds. As a result, the classification criterion adopted was to declare MR for individual or prototypical systolic intervals if the maximum amplitude peak was found to be at least 20 ms prior to S2. In the case of the configurations BBT, SBBT, BBTF, SBBTF

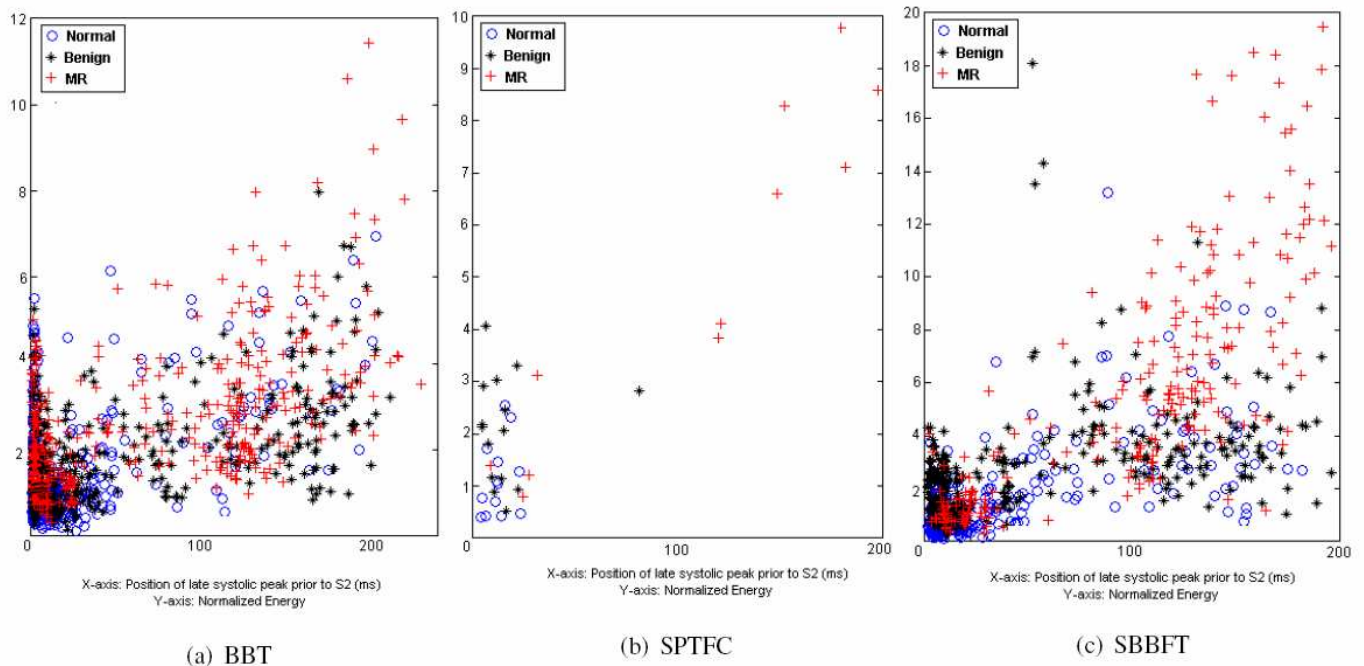


Fig. 6. Distribution of sample points across the feature space. In (a) and (c), each data point represents an individual beat. In (b), prototypical beats are constructed by merging together individual beats. This reduces the number of data points seen in (b) relative to (a) and (c).

and SBBTFC of Figure 5, where individual intervals were analyzed, subjects were classified as having MR if the majority of the intervals were labeled as showing MR. For SPTFC and SPTFC, where prototypical intervals were used, the classification of the subject was the same as that of the prototypical interval.

The effects of interval selection and time-frequency decomposition were determined by comparing the SBBT, BBTF and SBBTF configurations with BBT, while the clustering and prototypical approaches were evaluated by examining the SBBTFC, SPTFC and SPTFC configurations relative to SBBTF. The impact of using all processing steps can be assessed by the change in classification accuracy between the BBT and SPTFC configurations.

The results in Section 5 indicate that this fully automated classifier performs reasonably well compared to conventional approaches to computer-based auscultation, even those that are not fully-automated. However, there is still room for improvement, and we do not consider the overall performance of the classifier the central component of this work.

V. EXPERIMENTAL RESULTS

False negatives and false positives for various configurations of the MAAS Framework processing blocks (Figures 5(a)-5(g)) are shown in Table 1.

The SPTFC configuration makes full use of all processing functionality provided by the MAAS Framework. The effect of this approach, compared to the BBT configuration, which only makes use of the interval segmentation processing block to extract systolic segments, is to reduce false negatives from 3 to

1 (67% change) and false positives from 7 to 3 (56% change). Overall, the added processing significantly improves the accuracy of auscultation.

The constraint that subjects are classified under the BBT configuration based on the majority of the systolic intervals implicitly introduces a parameter of 50% into the classification scheme. This value could be changed to obtain better performance along either the false positive or false negative dimension. To obtain the same number of false negatives as the SPTFC configuration, a threshold of less than 26% must be chosen. This causes false positives to increase to at least 13. Conversely, to obtain the same number of false positives as the SPTFC configuration, a threshold of more than 70% must be chosen. This leads to at least 5 false negatives. Table 2 illustrates this effect. The performance of the BBT configuration is strictly worse than SPTFC, irrespective of the percentage of intervals we require to show MR before a patient is classified as having abnormal heart function.

Figures 6(a) and 6(b) provide a graphical representation of this behavior. Under the BBT configuration there is no clear delineation of the feature space into MR and non-MR regions. This effect restricts the accuracy of BBT classification. In the case of SPTFC, normal and benign data points collapse to a small region on the lower left, while MR data points are increasingly displaced from the origin.

In all experiments, passing the segmented intervals through a filter to selectively admit chunks of the raw audio with good SNR improved performance both in terms of false negatives and false positives. In the comparison between the BBT and SBBT configurations, false negatives decreased from 3 to 2, while in the case of BBTF and SBBTF, false positives improved from 11 to 9. This effect was largely the result of

being able to preferentially pick sections of the audio with improved signal to noise ratio for analysis. This was achieved through a two-fold process. The issue of variation across beats was resolved by means of physiological criteria to pinpoint intervals where heart sounds were likely to be loudest, while noisy stretches were simultaneously discarded. Pathological heart activity and normal heart function could both be heard more clearly as a consequence, leading to increased classification accuracy.

The use of time-frequency techniques to expose energy distribution across frequency scales for individual systolic segments decreased false negatives but led to an increased incidence of false positives. In the case of BBT and BBTF, false negatives decreased from 3 to 2, but false positives increased from 7 to 11. For SBBT and SBBTF, false negatives remained unchanged, while false positives increased from 7 to 9. The decrease in false negatives can be attributed to the increased visibility of energy caused by regurgitant murmurs at higher frequencies. Owing to the low volume of the heart sounds associated with cardiac disease, they are often washed out by the baseline S1-S2 signal and difficult to discern in the raw audio. Time-frequency analysis teases out these pathological signatures by exposing trends at higher frequency scales separate from the louder low frequency signals. The increase in false positives was caused by the presence of noise in our dataset, which resulted in increased energy at higher frequency scales.

Graphically, the combined effect of interval selection and time-frequency analysis can be seen in Figure 6(c). Compared to Figure 6(a), the density of MR data points in the lower left of the feature space decreases as they radiate outwards from the origin. On the other hand, data points corresponding to normal and benign subjects are clustered more heavily at lower values of both features.

The prototypical interval calculation stage addresses the false positive problem by retaining persistent spectral trends associated with MR while discarding random bursts of salt-and-pepper noise [24]. When added to the SBBTF configuration, there was a significant decrease in false positives from 9 to 3. False negatives increased from 2 to 3 for SPTF, because of the borderline misclassification of a paralyzed, 80 year old subject with severe Tricuspid Regurgitation (TR) in addition to moderate MR. A late systolic peak was found at 19.57 ms prior to S2, which was close to the decision boundary of 20 ms.

The experiments with time-frequency decomposition and prototypical interval calculation suggests that in practice the noise corrupting acoustic cardiac signals often affects the frequency ranges corresponding to diseased heart sounds. An approach to bandpass filter the signals would not be sufficient to remove noise artifacts. Instead, a technique similar to the one used to compute prototypical systoles is more appropriate, since it exploits the fact that for the class of patients at which this work is aimed heart sounds are generally quite similar over a period of time, and rapid deviations across beats are

TABLE I
FALSE NEGATIVES AND FALSE POSITIVES FOR EVALUATED CONFIGURATIONS OF
MAAS FRAMEWORK PROCESSING BLOCKS

Configuration	False Negatives	False Positives
BBT	3/11	7/28
SBBT	2/11	7/28
BBTF	2/11	11/28
SBBTF	2/11	9/28
SBBTFC	2/11	9/28
SPTF	3/11	3/28
SPTFC	1/11	3/28

TABLE II
EFFECT ON BBT OF VARYING THE THRESHOLD PARAMETER OF SYSTOLES THAT
MUST SHOW MR FOR A SUBJECT TO BE CLASSIFIED AS HAVING THE DISEASE

Threshold Percentage	False Negatives	False Positives
50%	3/11	7/28
26%	1/11	13/28
70%	5/11	3/28

TABLE III
FALSE NEGATIVES AND FALSE POSITIVES FOR A HUMAN EXPERT ON A SUBSET OF
THE FILES IN TABLE I

	Cardiologist	Automated Classifier
50%	3/6	1/6
26%	2/6	1/6

most likely due to noise.

Lung sounds are a prime candidate for removal in this manner. The range of frequencies occupied by respiratory artifacts may extend from less than 100 Hz to over 3000 Hz [17], a significant overlap with the frequency ranges examined during auscultation. Separating lung and heart sounds is therefore difficult by bandpass filtering the signal. The prototypical interval calculation stage circumvents this difficulty by exploiting the lack of temporal correlation between lung and heart sounds to obtain a relatively noise-free representation of recurring cardiac audio activity.

Prototypical interval calculation may make it more challenging to recognize subjects for whom the pathological condition does not occur on most beats. The process of aggregating information across a number of intervals, some of which may not display any abnormal activity may weaken the disease signature in the prototypical signal. The clustering and cluster selection steps are essential to dealing with this issue. In our experiments, when the prototypical interval calculation step was used in conjunction with both clustering and cluster selection, the overall effect was to reduce false negatives from 3 in SPTF to 1 for SPTFC, while maintaining false positives at 3.

When employed without the prototypical beat, the clustering approach in SBBTFC did not perform any differently from SBBTF. This example represents a case where an analytical stage of auscultation yields an improvement only when used in conjunction with another.

A subset of the recordings was presented to a board certified

cardiologist. Table 3 shows the results of a human expert on these files. There were a total of 3 false negatives and 2 false positives. For 2 of the 3 false negatives, the patients would have been recommended for an echocardiogram on the basis of likely aortic stenosis. The remaining false negative corresponded to the patient missed by the SPTFC configuration (Table 1). In this case, no discernable murmur was present and the labeling was done on the basis of regurgitation in the echocardiogram.

VI. SUMMARY AND CONCLUSIONS

In this paper, we developed an analytical perspective on cardiac auscultation, describing the complex task of interpreting heart sounds as the interplay of a sequence of simpler processing blocks. We provided an evaluation of how these components affect the overall goal of interpreting heart sounds and detailed the co-dependent manner in which processing blocks function relate to specified physiological phenomena. We further proposed a set of new processing stages and suggested refinements to existing algorithms, most notably in the area of segmenting raw audio signals into systolic and diastolic intervals.

The MAAS Framework provides a set of tools for automatically segmenting the audio signal into intervals corresponding to important sections of the cardiac cycle, isolating intervals with maximum diagnostic content, analyzing the time evolution of spectral characteristics across different frequency scales, and capturing statistically significant features corresponding to recurrent pathological signals while providing isolation from noise. We evaluated the relative contribution of each analytical stage of auscultation in correctly identifying the presence of MR and addressing the issues commonly encountered by physicians:

- Segmentation of the signal into individual beats and the beats into systole and diastole is the first stage of the process. Our approach to segmentation is distinctive from the works described in Section 2 in that we use the T wave in the ECG to more precisely identify the region of the audio signal containing S2. By searching over a shorter interval, we hope to reduce the probability of incorrectly detecting an artifact as S2. We seek further robustness to noise by making use of a patient-specific systolic length estimate. In this sense, our work is similar to the segmentation technique proposed in [8], which locates the R wave in ECG and then uses a multi-layer perceptron to estimate systolic length based on the patient heart rate, gender and age. In contrast to [8], our approach to calculating a patient-specific measure of systole focuses on inferring patient-specific information directly from the acoustic data
- Interval selection, clustering and cluster selection helped mitigate the problem of variation across beats. Interval selection preferentially admitted sections of

the audio recording where heart sounds were prominent and signal to noise ratio was good. Clustering and cluster selection grouped similar intervals together to identify statistically significant sound patterns. Our use of clustering in cardiac auscultation appears to be novel; although clustering has been applied previously to signals such as ECG (e.g., [3]), we did not find evidence of it being used in this context. Similarly, our review of existing work did not uncover prior work on using physiological criteria to automatically select intervals.

- Prototypical intervals reduced the impact of beat-to-beat variation by pooling information across cardiac cycles to capture temporally persistent trends and create a representative biological signal. Another key contribution of this step was to deal with noise such as respiratory artifacts, which may span the same frequencies occupied by diseased sounds and cannot always be removed by means of simple bandpass filtering. We did not find any examples in the literature describing the construction of prototypical beats to represent acoustical cardiac signals.
- Time-frequency decomposition dealt with the issue of pathological sounds being much quieter than the recurrent systolic signal by separating out low amplitude content at higher frequencies.

Some of the processing steps such as interval selection, time-frequency decomposition and prototypical interval construction were found to make a difference in their own right. Others, such as clustering and cluster selection improved performance only when used in conjunction with other modules such as prototypical interval construction. Even in the case of analytical stages that independently affected the interpretation of heart sounds, performance improved further when the processing blocks were chained together. The cumulative improvement provided by the MAAS Framework was greater than the sum of its individual parts.

We presented a sample application in Section 4 that leveraged the MAAS Framework to detect MR. A specific classifier was provided for illustrative purposes. Although classification performance the main focus of this paper, the results in Section 5 indicate that this simple MR classifier compares favorably with the best published results for other methods [8]–[14], which have been evaluated on similar data. Moreover, it is able to perform completely automated classification, without input from cardiologists to segment the signal or manually identify beats for further analysis.

There is reason to believe that the results derived from a simple MR classifier will generalize to a broad group of classifiers and cardiac diseases. Many of the recent computer-based classification systems (e.g., [8]–[14]) make use of similar features. The results in Section 5 show that use of the MAAS Framework processing blocks transforms the feature space so that it is easier to partition into homogeneous classes. Applications operating in the same feature space would

therefore benefit from the increased separability of data points, irrespective of the specific machine learning or statistical techniques employed.

We hope that the MAAS Framework will appeal to a large segment of the medical community. Educators can benefit from the modular representation of auscultation, which imparts a sense of structure to the overall process, and allows for systematic focus on the skills associated with individual analytical stages. Clinicians can make use of modules aimed at reducing the complexity of the tasks involved in auscultation, by employing them as audio-visual diagnostic aids [18]. Researchers can employ the modular processing provided by our framework to develop more powerful, fully-automated auscultation applications with improved classification.

ACKNOWLEDGMENT

We thank Dr. Collin Stultz for helping us compare the performance of our system to a human specialist, and Elena Glassman and Ali Shoeb for their insightful comments on the manuscript.

REFERENCES

- [1] National Center for Health Statistics, "Fast Stats A to Z: Heart Disease."
- [2] E Schwammenthal, C Chen, F Benning, M Block, G Breithardt and RA Levine, "Dynamics of Mitral Regurgitant Flow and Orifice Area. Physiologic Application of the Proximal Flow Convergence Method: Clinical Data and Experimental Testing," *Circulation*, 1994.
- [3] R Boussejot and D Kreisler, "Waveform Recognition with 10,000 ECGs," *Computers in Cardiology*, 2000.
- [4] E Braunwald, D Zipes, P Libby and R Bonow, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, W.B. Saunders Company, 2004.
- [5] A Pease, "If the Heart Could Speak," *Pictures of the Future*, 2001.
- [6] S Mangione and L Nieman, "Cardiac Auscultatory of Internal Medicine and Family Practice Trainees: A Comparison of Diagnostic Proficiency," *JAMA*, 1997.
- [7] W Tang and T Khoshgoftaar, "Noise Identification with the K-Means Algorithm," *IEEE Intl. Conf. on Tools with Artificial Intelligence*, 2004.
- [8] D Barschdorff, U Femmer and E Trowitzsch, "Automatic Phonocardiogram Signal Analysis in Infants Based on Wavelet Transforms and Artificial Neural Networks," *Computers in Cardiology*, 1995.
- [9] T Reed, N Reed and P Fritzson, "The Analysis of Heart Sounds for Symptom Detection and Machine-Aided Diagnosis," *2nd Conference on Modeling and Simulation in Biology, Medicine and Biomedical Engineering*, 2001.
- [10] W Thompson, C Hayek, C Tuchinda, J Telford and J Lombardo, "Automated Cardiac Auscultation for Detection of Pathologic Heart Murmurs," *Pediatric Cardiology*, 2001.
- [11] S Rajan, R Doraiswami, R Stevenson and R Watrous, "Wavelet Based Bank of Correlators Approach for Phonocardiogram Signal Classification", *Proc. of the IEEE-SP Intl Symp. on Time-Frequency and Time-Scale Analysis*, 1998.
- [12] C DeGroff, S Bhatikar, J Hertzberg, R Shandas, L Valdez-Cruz and R Mahajan, "Artificial Neural Network Based Method of Screening Heart Murmurs in Children," *Circulation*, 2001.
- [13] T Nakamitsu, H Shino, T Kotani, K Yana, K Harada, J Sudoh, E Harasawa and H Itoh, "Detection and Classification of Systolic Murmur for Phonocardiogram Screening", *Proc. of the 18th Intl Conf. of the IEEE Eng. in Med. and Biol. Soc.*, 1996.
- [14] W Myint and B Dillard, "An Electronic Stethoscope with Diagnosis Capability," *Proceedings of the 33rd IEEE Southeastern Symposium on System Theory*, 2001.
- [15] Z Syed, "MIT Automated Auscultation System," Master's thesis, MIT, 2003.
- [16] S Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 1989.
- [17] H Pasterkamp, S Kraman and G Wodicka, "Respiratory Sounds, Advances Beyond the Stethoscope," *Am. J. Respir. Crit. Care Med.*, 1997.
- [18] Z Syed, D Leeds, D Curtis, F Nesta, RA Levine and J Guttag, "Computer-Assisted Diagnosis of Cardiac Disorders", *IEEE Intl. Symposium on Computer Based Medical Systems*, 2006.
- [19] J Fraden and M Neuman, "QRS Wave Detection," *Med Biol Eng Comput.*, 1980.
- [20] G Friesen, T Jannett, M Jadallah, S Yates, S Quint and H Nagle, "A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms," *IEEE Trans Biomed Eng.*, 1990.
- [21] I Daubechies, "Ten Lectures on Wavelets," *CBMS-NSF Regional Conference Series in Applied Mathematics*, 1992.
- [22] R Duda, P Hart and D Stork, *Pattern Classification*, Wiley-Interscience, 2000.
- [23] Z Syed and John Guttag, "Prototypical Representation of Biological Signals", in preparation.
- [24] J Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, 1990.
- [25] D Berndt and J Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," *KDD*, 1994.



Zeeshan Syed received the S.B. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA in 2003, and is currently working towards a Ph.D. in the Harvard-MIT Health Sciences and Technology Division between the Department of Electrical Engineering and Computer Science, MIT and Harvard Medical School.

He has conducted research at the Computer Science and Artificial Intelligence Laboratory (formerly Laboratory for Computer Science) at MIT since 2001, working on projects related to the analysis and modeling of physiological signals, tools for the structured discovery of diagnostic markers, prototypical representations of biological activity, and the efficient detection of multi-modal patterns in long-term data. His research interests include biomedical signal processing, machine learning, algorithms and computational biology.

Mr. Syed is a member of Sigma Xi, Tau Beta Pi and Eta Kappa Nu. He is also a recipient of the William A. Martin, Morris J. Levin Masterworks, and the Global Technovators awards at MIT.



Daniel Leeds received the S.B. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA in 2005 and 2006 respectively. He is currently working towards a Ph.D. degree at the Carnegie Mellon University Robotics Institute and at the Center for the Neural Basis of Cognition in Pittsburgh, PA.

He has developed several tools to analyze physiological signals for medical diagnoses and for biological research. He has created software to produce 2D visualizations of the distributions of patients in diagnostically-relevant high-dimensional feature spaces, expanding upon cardiac sound analysis at the MIT Computer Science and Artificial Intelligence Laboratory. He has explored the kinematics of human speech at the MIT Research Laboratory of Electronics, and has studied the statistical structure of zebra finch communication, at Rockefeller University in New York City. He also has contributed to a system for modeling social networks based on speech patterns, at the MIT Media Laboratory, and has worked on message authentication codes, at New York University. He currently studies efficient

encodings for sensory data incorporating insights from neuroscience. His research interests include signal processing, machine learning, and human cognition.

Mr. Leeds is a National Science Foundation graduate research fellow



Dorothy Curtis received the S.B in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, the M.A.S.C.S. degree from the Metropolitan College of Boston University, Boston, MA , and the M.B.A. from Simmons College Graduate School of Management, Boston, MA in 1973, 1980 and 2000 respectively.

She is currently on the research staff at MIT and works in the area of data-driven medicine. She has developed the SMART system for monitoring patients in the waiting area of an emergency department. She has also participated in research in networking, object-oriented databases, and portable computers. She recently developed a tool for learning auscultation.



Francesca Nesta received an M.D. from the University of Brescia School of Medicine, Brescia, Italy in 1988.

From 2001 to 2004 she was a clinical and research fellow in echocardiography at the Massachusetts General Hospital, Boston, MA. She is currently a resident in medicine at Boston Medical Center, Boston, MA. Selected publications include “Leaflet concavity: a rapid visual clue to the presence and mechanism of functional mitral regurgitation” (J. American Society Echocardiography, 2003), and “A new locus for autosomal dominant mitral valve prolapse on chromosome 13: clinical lessons from genetics” (Circulation, 2005). Her research interests include the genetics of mitral valve prolapse, ischemic mitral regurgitation, and heart failure.



Robert A. Levine is Professor of Medicine at Harvard Medical School, and received an M.D. degree from Harvard University in 1978 after studying physiology at Cambridge University as a Marshall Scholar. He pursued postdoctoral training in cardiac physiology and imaging at the Massachusetts General Hospital, and has been involved in collaborative bioengineering projects at the Georgia and Massachusetts Institutes of Technology.

He is a recognized investigator and mentor in the field of noninvasive cardiovascular physiology with a leadership role in developing new techniques for quantifying heart function and applying them to study disease mechanisms and guide new therapies. His proof of the saddle geometry of the mitral valve led to clarification of the diagnosis of mitral valve prolapse as a foundation for this publication on digital heart sound analysis.

Dr. Levine is the author of nearly 200 peer-reviewed publications, and holder of patents for mitral valve repair and a new method to quantify valvular regurgitation.



John Guttag is the Dugald C. Jackson Professor of Electrical Engineering and Computer Science and a member of the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology (MIT), Cambridge, MA. Professor Guttag received AB and SM degrees from Brown University and his doctorate from the University of Toronto. He was on the faculty of the University of Southern California from 1975-1978, and joined the MIT faculty in 1979. From 1993-1998, he served as Associate Department Head from Computer Science of MIT's Electrical Engineering and Computer Science Department, and served as Head of that department from 1999-2004. His current research interests include the application of

wireless networking in healthcare, the development of techniques for the early detection of and interventions to control epileptic seizures, and techniques for analyzing large amounts of physiological data to discover information that can be used to assist in improving the management of chronic illness.