Shape-based Retrieval of Heart Sounds for Disease Similarity Detection

Tanveer Syeda-Mahmood, Fei Wang

[1] IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120.
{stf,wangfe}@almaden.ibm.com

Abstract. Retrieval of similar heart sounds from a sound database has applications in physician training, diagnostic screening, and decision support. In this paper, we exploit a visual rendering of heart sounds and model the morphological variations of audio envelopes through a constrained non-rigid translation transform. Similar heart sounds are then retrieved by recovering the corresponding alignment transform using a variant of shape-based dynamic time warping. Results of similar heart sound retrieval are demonstrated for various diseases on a large database of heart sounds.

Keywords: Sound pattern analysis, audio retrieval, curve analysis, healthcare application.

1 Introduction

The field of pattern recognition is becoming increasingly applicable to new modalities. In this paper we explore the application of computer vision techniques to an important class of audio signals, namely, heart sounds. Heart auscultation, i.e., listening to the sounds produced by the heart, is a common practice in the screening of heart disease. Although different diseases produce characteristic sounds, forming a diagnosis based on sounds heard through a stethoscope is a skill that takes years to perfect. Numerous studies have shown that as much as 87% of patients referred to cardiologists for evaluation are as a result of false alarms [2]. Thus software diagnostic tools that aid physicians in their diagnosis of heart sounds are needed.

Auditory discrimination of heart sounds is inherently difficult as these sounds are faint and lie at the lower end of the audible frequency range [1]. Although there are tools to visually render the sounds [3], we recently observed that the visual representation of heart sounds actually brings out the differentiating characteristics of various diseases more readily than the audio signal. Figure 1 illustrates this by recording the visual appearance of the audio signal within a single heart beat duration, from patients with various diseases. As can be seen, different diseases show characteristically different shape patterns. Further, it is also easier to spot the similarity in the sounds across patients with similar diseases through their visual representations. Figure 2 illustrates this for different patients diagnosed with the same disease. From these examples, it appears plausible that the disease similarity can be inferred by developing a measure for capturing the visual similarity of audio signals.

In this paper we address the similarity retrieval of heart sounds using spatial representations of heart sounds. Specifically, we extract perceptual envelopes of audio shapes that capture the essential low frequency disease-specific information. We then model the morphological variations in the spatial envelopes from the same disease class, as a constrained non-rigid translation transform. Matching involves recovering the corresponding alignment transform using a new shape-based dynamic time warping algorithm. Results of heart sound retrieval are reported for various diseases on a large database of heart sounds. Our experiments demonstrate that our method outperforms state-of-the-art heart sound signal analysis methods when they are adapted for use in similar heart sound retrieval.

The paper makes several important contributions. It addresses, for the first time, the problem of similarity retrieval of heart sounds to aid physician decision support (a new use of content-based retrieval methods). Secondly, it demonstrates the advantages of visual analysis of heart sounds over conventional audio processing methods to enhance disease similarity detection. Further, it proposes the use of envelope curves of heart sounds as an adequate representation to discriminate between diseases. Finally, the audio envelop representation used here is invariant to a number of factors including heart rate differences in young and adults, number of heart beat samples, noise in recordings, etc.

The rest of the paper describes this approach in detail. In Section 2 we introduce the domain of heart sounds and the need for modeling non-rigid time deformations for detecting visual similarity of audio signals. In Section 3, we discuss related work. In Sections 4-5, we develop a general method of modeling the shape changes in audio signals corresponding to similar sounds and present a method of audio shape matching. In Section 6, we describe the various pre-processing steps for enabling the shape matching. Finally, in Section 7 we present results demonstrating the effectives of shape similarity measures for discovering audio similarity among heart sounds.

2. The heart sound matching problem

The Heart sounds are produced during the heart cycle, where events such as the motion of the valves and the movement of the blood generate vibrations [1]. In healthy adults, the first heart sound (S1) (‘lub’) and second heart sound (S2) (‘dub’) are produced by the closure of the AV valves and semi-lunar valves as shown in Figure 3. In addition to these normal sounds, a variety of other sounds may be present including S3 which may be heard at the beginning of the diastole, during the rapid filling of the ventricles and S4, which may be heard in late diastole during atrial contraction, regurgitation sounds, i.e. murmurs which are noise-like sounds that are heard between the two major heart sounds during systole or diastole, and adventitious sounds, or clicks.

Matching heart sounds is difficult as they are highly non-stationary signals of short duration. Further, the energy content of low-frequency vibrations is much higher than that of high-frequency vibrations. Thus much of the discriminating heart sound information is captured in their low frequency envelopes which must be extracted. Next the variations in heart beat must be taken into account. This requires

extraction of single heart beats from signals which are not always periodic (particularly for cardiac arrhythmias). Further, the amplitude variations due to recording levels and varying digital stethoscope qualities affect the matching. Finally, due to variability in the systolic and diastolic phases across patients, both advancement as well as preponement of cardiac audio phases (sounds S1 and S3) is possible.

Figure 4 illustrates matching problem for heart sounds using two patients both diagnosed with Atrial Septal Defect (ASD). Figure 4a and b show the original sounds. Due to the sampling rate differences, a single heart beat corresponds to 6172 samples in Figure 4a while there are 34542 samples within a single heart beat for the sound in Figure 4b. The two signals, nevertheless sound very similar. Direct comparison of the two signals hardly reveals any similarity as can be seen through direct superposition in Figure 4c. Once the signals are normalized in both time and amplitude by scaling using the sampling rate and the range of 0 to 1, the similarity can be seen under a shift as shown in Figure 4d. Using cross-correlation, we can recover the shift to roughly align them as shown in Figure 4e. From this figure, it is clear that although a single fixed shift can bring them into rough alignment, matching peaks require different amounts of shift in different regions. Furthermore, the direction of the shift could be in opposite directions in different phases to correspond to the cases where the systolic and diastolic phases are affected differently (due to the disease). Figure 4f illustrates this non-uniformity in the time shift for the signals from Figure 4d which are already normalized in time (marked by arrows in the figure). Thus any measure of similarity in sounds for the same disease class needs to model non-uniformity in the scaling of time.

3 Related Work

While we are not aware of work on similarity retrieval of heart sounds, there is considerable work in audio signal analysis and retrieval in general [10,16,12,15,11], and heart sound signal analysis[6,7,4], in particular. Further, there is also work on computer-aided diagnosis of heart sounds [19,20] although no commercial systems exist in practice. Classical methods of audio analysis divide the audio into short-time segments and extract spatio-temporal features such as the zero-crossing rate (ZCR) [10], average energy in short-time segments, hidden Markov model features[14], etc. Methods based on frequency-domain information of the audio signal often form the speech spectrogram and extract features such as Mel-frequency cepstral coefficients (MFCC) [12], subband spectral flux, harmonicity prominence, spectral roll-off [15], beat frequency estimation [11], multi-resolution wavelets etc. Often, the spectral representations are generated from a moving overlapping window analysis of the time domain signal.

The short-time signal analysis methods are not suitable for modeling heart sounds due to (a) their highly non-stationary nature, (b) variability in heart rate, (c) disease-specific variations in the relative placement of S2,S2,S3,and S4 phases, and (d) due to the selective importance that needs to be given to the lower frequencies of the spectrum. Further, the choice of window size is not clear for heart sounds. Often, the

segments picked are end points of S1 or S3 phases. However, for diseased cases, there is smearing of the S1, and S3 phases (eg. Mitral regurgitation). Further, additional sounds such as S2 and S4 may be present. In such cases, it is not clear how to appropriately extract a coherent segment for spatial-frequency analysis. The method of MFCC suffers from the same limitation due to overlapping window analysis. Further, MFCC is well-known to be sensitive to even small amounts of noise.

Automatic analysis of heart sounds has been attempted for abnormality detection within a single sound. Early work on phono-cardiograms found correlation between sounds and various heart defects [19]. However, the sounds were studied by isolating single heart beat duration. The predominant approach in heart sound similarity detection has been based on feature extraction and classification as is conventional for audio analysis. In [19], a neural network was trained on heart sounds to classify several valve-related disorders. However, the identification of disorders was based on classifying sound into intuitive groups such as normal, split sound, open snap, etc based on detection of S1 and S2 sounds and noting their separation. For many diseases, such as pericardial rub, the periods S1 and S2 are hardly distinct for detection and hence are not reliable for audio similarity detection. In [20] again, a neural net classifier was used in conjunction with wavelet processing for heart sound classification. However, the system was trained on a single sample for each disease using different heart beat cycles. Heart sound analysis using time-frequency representations has also been common including recent uses of MFCC to heart sounds [5].

In all these approaches, no explicit variational modeling across different disease instances was performed other than using bounds on the sizes of the S1 and S3 phases. Finally, most of these methods assume that a single period of the heart sounds can be obtained. However, in a long recording, particularly in the case of arrhythmias, finding periodic repetitions by simple correlation is insufficient.

The visually modeling of heart sounds is analogous to visual modeling of music spectrograms reported in [25] although the spectral features were different. Finally, several methods of curve matching have been developed in literature[23,24]. Methods exist to match two curves using curvature-scale space [24], and to dynamically warp the curves into one another using dynamic time warping [23]. The curve matching approaches, in general, are sensitive to spurious and missing data. This condition is quite likely in modeling disease instances as the S1 or S3 phases are prolonged. The dynamic time warping algorithms for curve matching have mostly used time order and time proximity constraints. Shape-based dynamic time warping using combined shape cues including angle of corners, and orientation of their bisectors in combination with proximity have not been proposed previously.

4 Representing heart sounds using perceptual envelopes

In our approach, we model the heart sound (within a single heart beat) through a perceptual envelope. Various algorithms are available in literature for envelop extraction from signals including homomorphic filtering [8]. While these algorithms

are less sensitive to noise-related fluctuations, they frequently extract the low frequency component of the signal rather than render a faithful approximation of the perceptual envelope as can be seen from Figure 5.

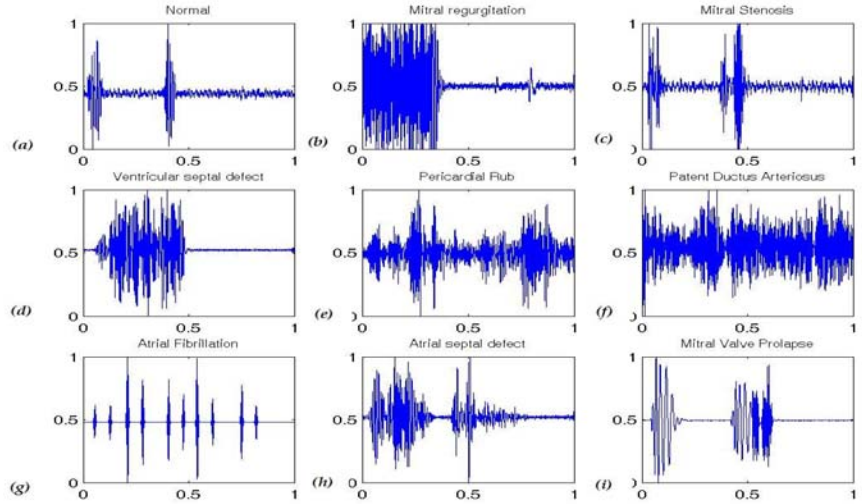

Figure 1. Illustration of different visual appearances for heart sounds for different diseases indicating visual discrimination of heart sounds is possible. (a)-(c) (d)-(f) , (g)-(i) left to right, top to bottom.

To extract audio envelope curve, we perform noise filtering using wavelet filters to remove the hissing noise that comes from digital stethoscopes. We then form a line segment approximation to the audio signal. This is a standard split-and-merge algorithm for line segment approximation that uses two thresholds, namely, a distance threshold δ and a line length threshold *l* to recursively partition the audio signal into a set of line segments. Each consecutive pair of line segments then defines a corner feature C_i. With δ=0.01 and *l*=5, a faithful rendering of the audio signal is made possible through a line segment approximation while retaining only 10% of the samples. The thresholds for curve parameterization are not as critical here as the shape matching algorithm we present next, is robust to missing and spurious features. We define the *audio envelope (AE)* as the set of points $AE = \{P_i\}$ *where* $P_i = C_j$ *for some* j *s.t.*

$P_i(y) \geq P_{i-1}(y) \,\&\, P_i(y) \geq P_{i+1}(y)$ and $P_i(y) \geq baseline$.

Only the values of the signal above the baseline are retained in the maxima envelope curve. Similarly, all peaks below the baseline for the minima envelope curve.

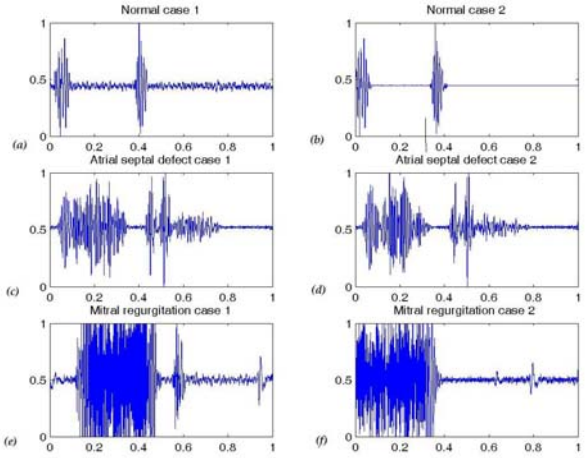

Figure 2. Illustration of similarity retrieval for heart sounds. The top matches in each case are matching patients with the same disease. Caption runs from (a)-(b), (c)-(d), (e)-(f) , Left to right, top to bottom in the figure.

Figure 3. Illustration of four major heart sounds.

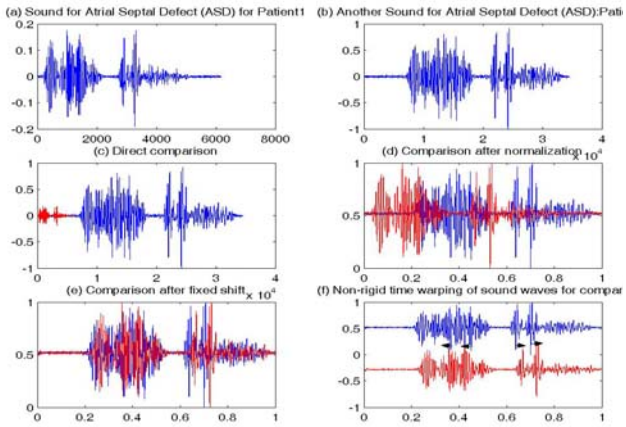

Figure 4. Illustration of the alignment problem for heart sounds.

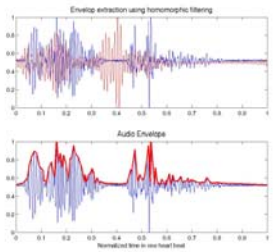

Figure 5. Illustration of envelope curve extraction of sounds. (a) Homomorphic filtering-based low frequency approximation (shown in red), (b) perceptual envelope.

Figure 5b illustrates the envelope curve extracted from a heart sound. As can be seen, the boundary shape of the signal is well-captured in the envelope curve. Once the envelope curve f(t) is extracted, its shape can be represented by the curvature change points or corners on the envelope curve. The shape information at each corner is captured using the following parameters

$$S(\vec{f}(t_i))=<t_i, \vec{f}(t_i), \theta(t_i), \varphi(t_i)> \quad (1)$$

These features were chosen to facilitate matching of audio envelopes. Using the angle of the corner ensures that wider complexes are not matched to narrow complex as these can change the disease interpretation. The angular bisector, on the other hand, ensures that polarity reversals such as inverted waves can be captured. Here $\theta(t_i)$ is the included angle in the corner at t_i, and $\varphi(t_i)$ is the orientation of the bisector at corner t_i.

5 Modeling morphological variations

We now develop a model of morphological variations in the shape of heart sounds envelopes for different instances of a disease.

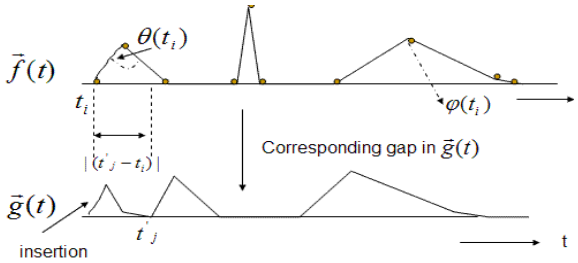

Figure 6. Illustration of the shape matching problem for envelop curves.

Referring to Figure 6, consider an envelop curve g(t) corresponding to a heart sound. Consider another curve f(t) that is a potential match to g(t), i.e. comes from a different patient diagnosed with the same disease. The curve f(t) is considered perceptually similar to g(t) if a non-rigid transform characterized by [a,b, Γ] can be found such that

$$| f'(t) - g(t) | \leq \delta \quad (2)$$

where | | represents the distance metric that measures the difference between f'(t) and g(t), the simplest being the Euclidean norm and

$$f'(t) = af(\Phi(t)) \quad \text{with} \quad \Phi(t) = bt + \Gamma(t) \quad (3)$$

where the (bt) is the linear component of the transform and Γ is the non-linear translation component. The parameters a and b can be recovered by normalizing in amplitude and time. We can normalize in amplitude by transforming f(t) and g(t) such that

$$\hat{f}(t) = \frac{f(t) - f_{\min}(t)}{f_{\max}(t) - f_{\min}(t)} \quad \text{and} \quad \hat{g}(t) = \frac{g(t) - g_{\min}(t)}{g_{\max}(t) - g_{\min}(t)} \tag{4}$$

so that a=1. To eliminate solving for b, we can normalize the time axis by dividing the heart rate. Suppose the sampling rate of points on the curve f(t) is F_S. Let a periodicity detection algorithm signal the heart rate period to be T_1 (Periodicity detection is discussed in Section 5). Then dividing by the heart period samples, we can ensure that that all time instants lie in the range [0,1]. Thus the time normalization can be easily achieved as:

$$\vec{f}(t) = \hat{f}(t/T_1) \quad \vec{g}(t) = \hat{g}(t/T_2) \tag{5}$$

where T1 and T2 are the heart beat durations of f(t) and g(t) respectively. With this time normalization, b=1. Such amplitude and time normalization automatically makes the shape modeling invariant to amplitude variations in audio recordings, as well as variations in heart rate across patients. Since the non-uniform translation Γ is a function of t, we can avoid computational overhead by recovering it at important fiducial points such as the corners, and recover the overall shape approximation by interpolation. Let there be K features extracted from $\vec{f}(t)$ as $F_K = \{(t_1, \vec{f}_1(t_1)), (t_2, \vec{f}_2(t_2)) \ldots (t_K, \vec{f}_K(t_K))\}$ at time $\{t_1, t_2, .. t_K\}$ respectively. Let there be M fiducial points extracted from $\vec{g}(t)$ as $G_M = \{(t'_1, \vec{g}_1(t'_1)), (t'_2, \vec{g}_2(t'_2)) \ldots (t'_M, \vec{g}_M(t'_M))\}$ at time $\{t'_1, t'_2, .. t'_M\}$ respectively. If we can find a set of N matching fiducial points $C_\Gamma = \{(t_i, t'_j)\}$, then the non-uniform translation transform Γ can be defined as:

$$\Gamma(t') = \begin{cases} t_i \quad if\, t = t'_j \, and \, (t_i, t'_j) \in C_\Gamma \\ t_r + \left(\frac{t_s - t_r}{t'_l - t'_k} \right) (t' - t'_k) \quad where\, (t_r, t'_k), (t_s, t'_l) \in C_\Gamma \end{cases} \tag{6}$$

and t'_k is the highest of {tj'} $\leq t'$ and t'_l is the lowest of {tj'} $\geq t'$ that have a valid mapping in C_Γ. Other interpolation methods besides linear (eg. spline) are also possible.

Using Equations 5 and 6, the shape approximation error between the two curves is then given by:

$$| f'(t) - g(t) | = | \vec{f}(t' = \Gamma(t')) - g(t') | \tag{7}$$

For each g(t), we would like to select Γ such that it minimizes the approximation error in (7) while maximizing the size of match C_Γ. Finding the best matching audio based on shape can then be formulated as finding the g(t) such that

$$g_{best} = \arg\min_{g} | \vec{f}(\Gamma(t^{'}) - \vec{g}(t^{'}) |$$ (8)

while choosing the best Γ for each respective candidate match g(t).

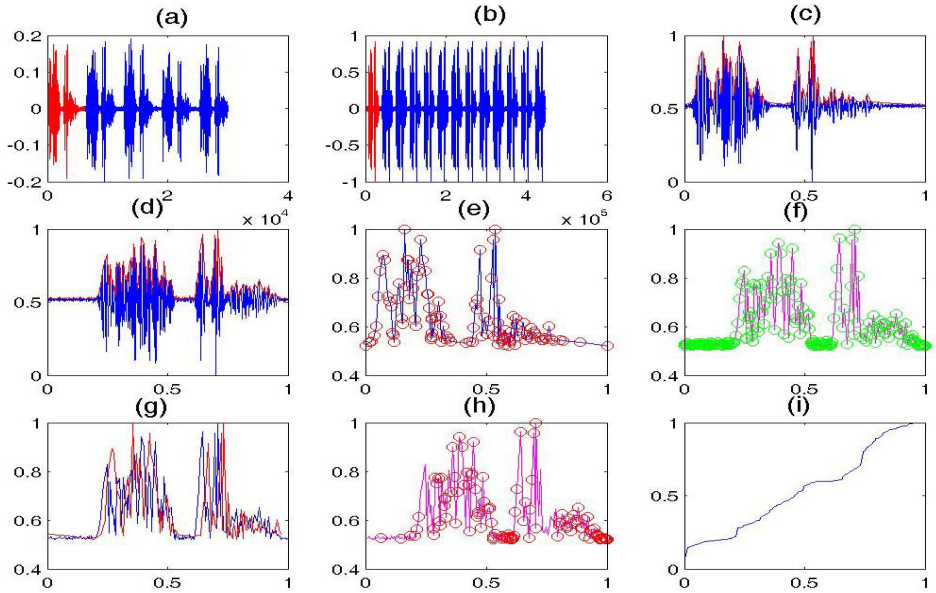

Figure 7. Illustration of shape-based dynamic time warping.

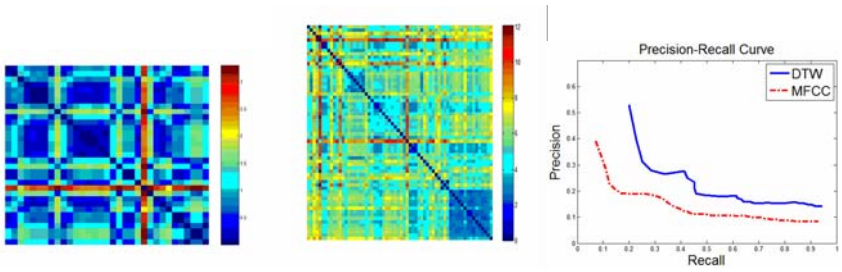

Figure 8. Illustration of the comparison between shape-based time warping and MFCC algorithm for pair-wise discrimination. (a) Confusion matrix using our algorithm (b) Confusion matrix using MFCC.

5.1 Shape-based dynamic time warping

If we consider the feature set F_K, G_M extracted from the respective curves as sequences, the problem of computing the best Γ reduces to finding the best global subsequence alignment using the dynamic programming principle. The best global alignment maximizes the match of the curve fragments while allowing for possible gaps and

insertions. Gaps and insertions correspond to signal fragments from feature set F_K that don't find a match in set G_M and vice versa. In fact, the alignment can be computed using a dynamic programming matrix H where the element H(i,j) is the cost of matching up to the ith and jth element in the respective sequences. As more features find a match, we want the cost to increase as little as possible. The dynamic programming step in our case becomes:

$$H_{i,j} = \min\begin{cases} H_{i-1,j-1} + d(\vec{f}(t_i), \vec{g}(t'_j)) \\ H_{i-1,j} + d(\vec{f}(t_i), 0) \\ H_{i,j-1} + d(0, \vec{g}(t'_j)) \end{cases} \tag{9}$$

with initialization as $H_{0,0} = 0$ and $H_{0,j} = \infty$ and $H_{i,0} = \infty$ for all 0< i <= K, and 0<j<=M. Here d(.) is the cost of matching the individual features described next. Also, the first term represents the cost of matching the feature point $\vec{f}(t_i)$ to feature point $\vec{g}(t'_j)$ which is low if the features are similar. The second term represents the choice where no match is assigned to feature $\vec{f}(t_i)$.

5.2 Shape similarity of envelop curves

After the transformation Γ is recovered, the similarity between two envelop curves is given by the cost function $d(\vec{f}(t_i), \vec{g}(t'_j))$

$$d(\vec{f}(t_i), \vec{g}(t_j) = \begin{cases} \sqrt{(t_i - t'_j)^2 + (\vec{f}(t_i) - \vec{g}(t'_j)^2 + (\theta(t_i) - \theta(t'_j))^2 + (\varphi(t_i) - \varphi(t'_j))^2} & \text{if } \begin{matrix} |(t_i - t'_j)| \le \lambda_1 \\ |(\vec{f}(t_i) - \vec{g}(t'_j)| \le \lambda_2 \\ |(\theta(t_i) - \theta(t'_j))| \le \lambda_3 \\ |(\varphi(t_i) - \varphi(t'_j))| \le \lambda_4 \end{matrix} \\ \infty \quad \text{otherwise} \end{cases} \tag{10}$$

The thresholds $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ are determined through a training phase in which the expected variations per disease class are noted. The cost function $d(\vec{f}(t_i), 0)$ can be computed by substituting $t'_j = 0,\ \vec{g}(t'_j) = 0,$ and $\theta(t'_j) = 0, \varphi(t'_j) = 0$ in Equation 10. The cost function $d(0, \vec{g}(t'_j))$ can be similarly computed.

Thus using this approach two heart sounds are considered similar if enough number of fiducial points between query and target envelop curves can be matched using shape-based dynamic time warping. In general, due to the period estimation offset errors, the signals may have to be circularly shifted by a fixed translation for a rough alignment before the fine non-rigid alignment described above.

6. Feature pre-processing

The above formulation assumed that a single heart beat duration signal could be isolated prior to initiating shape matching. We now describe the pre-processing steps that lead to such an isolation.

6.1 Periodicity detection

While normal heart sounds show good repetition, the periodicity is surprisingly difficult to spot for abnormal heart sounds where the repetitions can be nested or irregular (arrhythmias). Simple auto-correlation is often insufficient for this purpose. A robust periodicity detector was, therefore, developed that treats periodicity detection as the problem of recovering a shift/translation that best aligns two self-similar curves. Consider a periodic curve g(t) with period T. Then by definition g(t) = g(t+kT) for all multiples k=0,1,2,…. Consider a candidate period τ. Form a curve f(t) by shifting g(t) by τ, i.e.

$$f(t) = \begin{cases} g(t-\tau) & if\ t \geq \tau \\ g(t) & otherwise \end{cases} \tag{11}$$

Then define a function $R(\tau)$ that records the number of curve features that can be verified to satisfy the periodicity condition based on the current estimate of the period as

$$R(\tau) = \frac{|\{g(t_i)\}|}{\max\{N-\tau, \tau\}} \quad s.t. |\, g(t_i) - f(t)_i \,| \leq \varepsilon \tag{12}$$

where N is the total number of points on the curve. The above function can be computed in linear time in comparison to the quadratic time for the autocorrelation function. The function $R(\tau)$ shows peaks at precisely those shifts which correspond to periodic repetitions of the curve. The most likely period is then taken as the smallest τ with the most integer multiples in the allowed range of heart beats (40-180 beats/minute). In general, there may be more than one candidate for period, particularly when the periodic repetitions are nested. Our algorithm finds all such overlapping repetitions and tests the dynamic time warping algorithm with each such choice.

7. Results

The audio shape matching algorithm was tested on a large database of heart sounds. The data came from various teaching hospitals and reference CDs provided by digital stethoscope makers (Littman). Thus the ground truth labels for diseases were known during the evaluation. Currently, the collection has over several hundred

heart sound examples for various kinds of murmurs, Mitral regurgitation, Mitral Stenosis, septal defects, Cardiomyopathy, etc. Each disease was represented by at least 6-10 patient samples in the database.

Figure 7 illustrates the steps involved in audio shape matching through a sample retrieval. Figure 7a shows the query audio signal and Figure 7b shows a matching audio sound from the database. The result of periodicity detection is shown overlaid in the respective figures in red. Figure 7c and d show the audio envelopes extracted from the respective signals normalized in amplitude and time over a single heart beat duration. Figure 7e and 7f show fiducial features extracted for matching from the query and database signals. Direct alignment of the two signals by cross-correlation is shown in Figure 7g. As can be seen, alignment errors are still present. The shape-based dynamic time warping alignment is indicated in Figure 7i. The projection of fiducial features from the query (green circle) onto the matching database curve (in magenta) using the non-rigid alignment is indicated in Figure 7h. In this example, 110 of the 114 of query fiducial features from Figure 7e have found a match with fiducial features of the database curve. From this figure, it is clear that although a single fixed shift can bring them into rough alignment, matching the signals requires different amounts of shift in different regions. Furthermore, the direction of the shift could be opposite in different sections when the systolic and diastolic phases are affected differently due to the disease. Figure 2a-f illustrates three other examples of heart sound similarity retrieval. In each case, the query is shown on the left and the best matching heart sound is shown on the right.

7.1 Comparison with MFCC

We compared the performance of our algorithm with a traditional signal processing approach using Mel-frequency Cepstral Coefficients (MFCC) [2]. The MFCC method was chosen as it has been the most successful of the audio analysis approaches and has recently been used to classify heart diseases [5]. We implemented a version of MFCC in which we divided the raw audio signals (no periodicity detection) into short-time segments of 400 samples and MFCC coefficients were then extracted from each segment using the short-time Fourier transform (STFT). The coefficients of all segments were used to form the feature vector. Euclidean distance was used to find nearest matches.

We tested the two techniques for grouping sounds from similar diseases and discriminating sounds from different diseases using the pair-wise similarity matrix. Thus a technique that discriminates well would form bands along the diagonal in the similarity matrix. Figure 8a and 8b show the similarity matrices using our algorithm and using MFCC. As can be seen, the banding structure is clearer using our method in comparison to MFCC where the banding is evident mostly for normal heart sounds. Further, the outlier (the orange colored signal) that has no match in the database clearly stands out using the DTW alignment.

7.2 Precision recall performance

Finally, we evaluated the precision and recall of the two methods by using all available samples per disease as queries and retrieving matches from the database for various choices of K and retained top K matches. Precision and recall were defined as

$$\text{Re}\,call = \frac{\#\ correct\ matches\ selected}{\#\ of\ \ correct\ matches\ \ present} \qquad \text{Pr}\,ecision = 1 - \frac{\#\ incorrect\ matches\ selected}{\#\ \ of\ \ matches\ returned}$$

The precision and recall values were averaged over the queries tested for the respective disease classes. Figure 8c shows the performance of the two methods on the entire audio database. Again, as can be seen, both precision and recall are higher using the more precise audio shape matching algorithm over MFCC. Due to the averaging over a large number of queries, the precision values are somewhat lower than expected here. Sample precision and recall values for some diseases averaged over all queries of the same disease class are shown in Table 1 below indicating actually good precision and recall for specific diseases. In general, we observed that the top K matches returned by our method often contained signals whose visual appearance were similar and the corresponding auscultation sounds heard were similar.

Table 1. Illustration of precision and recall.

Disease	• DTW Precision	• DTW Recall	• MFCC Precision	• MFCC Recall
Mitral regurgitation	• 56.67%	• 70.83%	• 43.33%	• 50.0%
Patent Ductus Arteriosus	• 100%	• 75%	• 50.0%	• 37.5%
Pericardial Rub	• 67.2%	• 85.6%	• 54.4%	• 45.7%
Mitral stenosis	• 76.2%	• 78.3%	• 45.2	• 39.8%
Ventricular septal defect	• 65.4%	• 82.3%	• 54.2%	• 48.9%
Atrial Fibrillation (AF)	• 78.2%	• 84.5%	• 64.7%	• 54.3%

8. Conclusions

In this paper, we have presented a novel algorithm for shape based retrieval of heart sounds. Unlike existing work based on feature extraction and classification, we take the approach of nonrigid shape alignment for retrieval. The algorithm is independent of disease specifics and is potentially applicable to other audio signals where their visual appearance is sufficiently discriminatory. Our experiments demonstrate that this approach significantly outperforms the current state-of-the-art approached based on heart sound signal analysis.

References

[1] Erikson B., Heart sounds and murmurs: A practical guide. Mosby-Year, 1997, 9-12.

[2] A.F. Pease “If the heart could speak” , online reference from.w4.siemens.de/FuI/en/archiv/pof/heft2_01/artikel19/index.html.
[3] http://www.zargis.com/WhitePaperIntro.asp
[4] See survey of papers on auscultation analysis at http://www.bsignetics.com/bioengineeringpapers.htm.
[5] I. Kamarulafizam et al.,”Heart Sound Analysis Using MFCC and Time Frequency Distribution,” in Proc. 3rd Intl. Conf. on Biomedical Eng., Kaula Lampur, 2006.
[6] O.A. Alim, N. Hamdy, and M.A. El-Hanjouri, “Heart diseases diagnosis using heart sounds,” in Proc. National Radio Science Conference (NRSC), pp. 634-640, 2002.
[7] Hebden, J.E.; Torry, J.N. Identification of aortic stenosis and mitral regurgitation by heart sound analysis. Computers in Cardiology 1997, 7-10 Sept. 1997 Pages:109 - 112.
[8] I. Rezek, and S.J. Roberts, “Envelope Extraction via Complex Homomorphic Filtering,” Research Report TR-98-9, June 1998.
[9] Univ. of Washington Medical School (http://depts.washington.edu/physdx/index.htm).
[10] L. Rabiner and R. Shafer, “Digital Processing of Speech Signals”, Prentice Hall: NJ.
[11] Jonathan Foote, Matthew Cooper. Audio Retrieval by Rhythmic Similarity. Proceedings of ISMIR 2002, Paris, France, October 2002.
[12] J. T. Foote, “Content-based retrieval of music and audio,” Proc of SPIE, vol.3229, pp.138-147, 1997.
[13] L. Lu, H.-J. Zhang, H. Jiang, "Content Analysis for Audio Classification and Segmentation", IEEE Trans. on Speech and Audio Processing, Vol.10, No.7, pp.504-516, 2002.
[14] Lie Lu, Rui Cai, and Alan Hanjalic. “Towards A Unified Framework for Content-based Audio Analysis”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2005, Vol. II, pp1069-1072, 2005.
[15] R. CAi, L. Lu, H.-J. Zhang, Using structure patterns of temporal and spectral feature in audio similarity measure, Proc. ACM Multimedia 2003., 219-222.
[16] Cai, R., Lu, L., Hanjalic, A., Zhang, H.-J., and Cai, L.-H. A flexible framework for key audio effects detection and auditory context inference. in IEEE Trans. Speech Audio Processing, May, 2006.
[17] Guodong Guo and Stan Z. Li, Content-Based Audio Classification and Retrieval by Support Vector Machines, IEEE Transactions on Neural Networks, vol.14, no.1, Jan. 2003.
[18] Jandre, F.C.; Souza, M.N, Wavelet analysis of phonocardiograms: differences between normal and abnormal heart sounds Proceedings of the 19th Annual International Conference of the IEEEVolume 4, Issue , 30 Oct-2 Nov 1997 Page(s):1642 - 1644 .
[19] O.A.Alim et al, “Heart diseases diagnosis using heart sounds,” in Proc. 19th National Radio Science Conference, Alexandria, March 2002.
[20] Say, O.; Dokur, Z.; Olmez, T. Classification of heart sounds by using wavelet transform, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] EMBS/BMES Conference, Volume 1, Issue , 2002 Page(s): 128 – 129.
[21] Iead Rezek and Stephen J. Roberts (1998). Envelope Extraction via Complex Homomorphic Filtering. Research Report TR-98-9, June 1998.
[22] Elfeky, M.G. Aref, W.G. Elmagarmid, A.K . WARP: time warping for periodicity detection. In Fifth Intl. Conf. on Data Mining, p. 8, Nov. 2005.
[23] H.J Wolfson, “On curve matching,” in IEEE Trans. PAMI, pp.483-489, vol.12, May 1990.
[24] B. Avants and J. Gee, “Continuous curve matching with scale-space curvature and extrema-based scale selection, “ in Proc. Scale-space Methods in Computer Vision, p.1079, 2003.
[25] Y. Ke, D. Hoiem, and R. Sukthankar. “Computer vision for music identification,” in Proc. CVPR 2000.