Unsupervised Classification of Heart Sound Recordings

Wei-Ho Tsai*, Sung-How Su† and Cin-Hao Ma*
*Department of Electronic Engineering & Graduate Institute of computer and Communication Engineering, National Taipei University of Technology, Taipei, Taiwan, R.O.C.
E-mail: {whtsai, t101419012}@ntut.edu.tw Tel: +886-2-27712171 Ext. 2257
†Pojen General Hospital, Taipei, Taiwan, R.O.C.
E-mail: shscvs@mac.com

Abstract—**An unsupervised framework for classifying heart sound data is proposed in this paper. Our goal is to cluster unknown heart sound recordings, such that each cluster contains sound recordings belonging to the same heart diseases or normal heart beat category. This framework is more flexible than the existing supervised classification of heart sounds by the case when heart sound data belong to undefined categories or when there is no prior template data for building a heart sound classifier. To this end, methods are proposed for heart sound feature extraction, similarity computation, cluster generation, and estimation of the optimal number of clusters. Our experiments show that the resulting clusters based on our system are roughly consistent with the heart beat categories defined by human labeling.**

I. INTRODUCTION

Audiovisual data has been the lion's share of data circulated on the Internet, and hence the techniques for helping people retrieve audiovisual data is gaining importance. In particular, compared to image and video, audio data is hard to browse, making the tremendous time consumption for manual indexing. Thus, it is highly desirable to develop automatic methods to analyze and organize various audio data, in order to lessen or replace the human efforts for listening to and then labeling them. In this work, we study the problem of classifying heart sound recordings for cardiac auscultation database indexing. This research is motivated by the increasing trend on sharing biomedical data on the Internet for academic purpose. Thank to the advent of electronic stethscope, auscultation can be fully recorded, documented, and shared; hence, organizing heart sound data efficiently is of necessity. Compared to massive amounts of studies on speech and music data classification, research on heart sound data classification has not yet been explored seriously.

Heart sounds are the noises generated by blood turbulence through the closing heart valves. In healthy adults, heart sounds come in two types: the first heart sound (s1), caused by the closure of the atrioventricular valves, and the second heart sound (s2), caused by the closure of the semilunar valves. In children or teenagers, it is common and normal to hear the third heart sound (s3), caused by the oscillation of blood back and forth between the walls of the ventricles. In elderly people, there might be the fourth heart sound (s4), caused by the vibration of the ventricular wall during atrial contraction. s4 is proved to be a sign or symptom of heart failure during diastolic period. In general, the frequency of s1 is lower than that of s2, and the duration of s1 is longer than that of s2. As to s3 and s4, the former occurs from 0.1 to 0.2 second after s2, while the latter occurs from 0.07 to 0.1 second before s1. Both of them are low in frequency. In addition to s1, s2, s3, and s4, there are numerous heart murmurs arising mainly from heart problems or diseases. Murmurs range from very faint to very loud. The number of categories of murmurs may be regarded as infinity.

To date, there is only limited research focusing on identifying heart sound data automatically. In [1], the authors used instantaneous energy and frequency estimations to identify six heart beat categories: 1) normal first (sl) and second heart sound (s2), 2) mitral regurgitation with third heart sound, 3) widely splitting of s2, 4) mitral stenosis, 5) mitral stenosis - mitral regurgitation, 6) fourth heart sound (s4) gallop. In [2], a heart sound identifier based on neural network was developed. The authors studied the identification of the three heat beat categories: 1) normal, 2) with holosystolic murmur, 3) with midsystolic murmur. In [3], the authors used hidden Markov model to identify heart sounds. In [4], a method based on wavelet transform together with incremental self-organizing map was developed to identify 14 heart beat categories. In addition, the work in [5] studies the use of short-time Fourier transform and support vector machines to identify 40 heart beat categories. In [6], the heart sound identification is extended to a problem of heart sound retrieval, in which heart sound data are converted from waveform representations into sequences of MIDI (Musical Instrument Digital Interface) notes. In [7,8], computer-assistance software is designed to help doctors diagnose heart diseases

It is worth noting that all of the related studies mentioned above are based on a supervised classification framework, which relies on manually-labeled heart sound data to build a classifier for determining the attributes of unknown heart sounds. However, as labeling heart sound data manually is very time-consuming, it is not easy to scale up the classifier to add in more categories of heart beats or handle undefined categories. Instead of the supervised framework, this work proposes an unsupervised framework for classifying sound data. Given a set of N unknown heart sound recordings, each belongs to one of the P heart beat categories, where $N \geq P$, and P is also unknown, our aim is to produce a partitioning of the N heart sound recordings into M clusters such that $M = P$, and each cluster consists exclusively of recordings belongs to only one heart beat category. As there is no need of labeling heart sound data manually, the proposed clustering framework is not limited to handling the heart sounds of human-defined categories. To the best of our knowledge, no prior work has been done to investigate this problem.

II. System Configuration

The proposed system for clustering heart sound recordings consists of four major components: heart sound characterization, computation of inter-recording similarities, generation of clusters, and estimation of the number of involved heart beat categories. The heart sound characterization aims to extract useful features from heart sound data. The similarity computation is designed to produce larger values for similarities between heart sound recordings from the same categories and smaller values for similarities between heart sound recordings from different categories. Then, clusters are generated in a bottom-up agglomerative manner, which sequentially merges the sound recordings deemed similar to each other. The outcome of the agglomeration procedure is a cluster tree with the number of clusters ranging from 1 to N. The tree is then cut by determining the optimal number of clusters, which corresponds to the number of heart beat categories.

Depending on the application, there are a number of ways to evaluate the performance of the heart sound recording clustering. In this study, we use two metrics: cluster purity [9,11], and the Rand index [10]. Cluster purity indicates the degree of correct clustering. It is represented by the probability that if we pick any recording from a cluster twice at random, with replacement, both of the selected recordings are from the same heart beat categories. Specifically, the purity of cluster c_m is computed by

$$\rho_m = \sum_{p=1}^{P} \left(\frac{n_{mp}}{n_{m*}} \right)^2, \quad (1)$$

where n_{m*} is the total number of sound recordings in cluster c_m, and n_{mp} is the number of sound recordings in cluster c_m that are from the p-th category. From Eq. (1), it follows that $n^{-1}_{m*} \leq \rho_m \leq 1$, in which the upper bound and lower bound reflect that all the within-cluster sound recordings are from the same categories and completely different categories, respectively. To evaluate the overall performance of M-clustering, we compute an average purity

$$\bar{\rho} = \frac{1}{N} \sum_{m=1}^{M} n_{m*} \, \rho_m . \quad (2)$$

Obviously, a perfect clustering should produce an average purity of one. However, this does not work both ways. The value of the average purity generally increases as the number of clusters increases, since the metric does not consider errors that place sound recordings belonging to the same heart beat category in different clusters. Hence, the cluster purity is only suitable for comparing the performance of different clustering methods if the number of clusters is specified *a priori*.

In contrast, the Rand index indicates the degree of incorrect clustering. It is defined as the number of sound recording pairs from the same category that are placed in different clusters, or sound recording pairs from different categories that are placed in the same cluster, i.e.,

$$R(M) = \sum_{m=1}^{M} n_{m*}^2 + \sum_{p=1}^{P} n_{*p}^2 - 2 \sum_{m=1}^{M} \sum_{p=1}^{P} n_{mp}^2 , \quad (3)$$

where n_{*p} is the number of sound recordings from the p-th category. Obviously, the smaller the value of $R(M)$, the better the clustering performance will be. The Rand index can be alternatively represented as a clustering error rate:

$$E(M) = \frac{\sum_{m=1}^{M} n_{m*}^2 + \sum_{p=1}^{P} n_{*p}^2 - 2 \sum_{m=1}^{M} \sum_{p=1}^{P} n_{mp}^2}{\sum_{m=1}^{M} n_{m*}^2 + \sum_{p=1}^{P} n_{*p}^2} \times 100\,\%. \quad (4)$$

Unlike the cluster purity metric, which favors a large M value, the Rand index usually decreases with an increase in the value of M initially, and reaches the minimum at $M = P$. When $M > P$, the Rand index starts to increase as the value of M increases. Hence, the Rand index is suitable for comparing the performance of different clustering methods involving different numbers of clusters generated.

III. Heart Sound Characterization

Because heart sounds are similar to drum sounds, we can characterize them by means of musical notes. An intuitive way to represent musical notes is the use of MIDI number. Specifically, each heart sound recording is converted from its waveform representation into a sequence of MIDI note numbers. Let e_k, $1 \leq k \leq K$, be the inventory of possible notes involving in heart sounds. Our aim is to determine which among the K possible notes is most likely produced at each instant in a heart sound recording. We apply the strategy in [6] to solve this problem. First, the heart sound recording is divided into frames by using a B-length sliding Hamming window, with $0.5B$-length overlapping between frames. Every frame then undergoes a Fast Fourier Transform (FFT) with size J. Let $x_{t,j}$ denote the signal's energy with respect to FFT index j in frame t, where $1 \leq j \leq J$, and $x_{t,j}$ has been normalized to the range between 0 and 1. Then, the signal's energy on k-th note in frame t can be estimated by

$$\hat{x}_{t,k} = \max_{\forall j, U(j)=e_k} x_{t,j} , \quad (5)$$

and

$$U(j) = \left\lfloor 12 \cdot \log_2 \left(\frac{F(j)}{440} \right) + 69.5 \right\rfloor, \quad (6)$$

where $\lfloor\ \rfloor$ is a floor operator, $F(j)$ is the corresponding frequency of FFT index j, and $U(\cdot)$ represents a conversion between the FFT indices and the MIDI note numbers.

Ideally, if note e_k occurs in frame t, the resulting energy, $\hat{x}_{t,k}$, should be the maximum among $\hat{x}_{t,1}, \hat{x}_{t,2}, ..., \hat{x}_{t,K}$. However, it is sometimes the case that the energy of a true note is smaller than that of its harmonic note. To avoid the interference of harmonics in the estimation of true notes, we use the strategy of Sub-Harmonic Summation (SHS) [12], which computes a value for the "strength" of each possible note by summing the signal's energy on a note and its harmonic note numbers. Specifically, the strength of note e_k in frame t is computed using

$$y_{t,k} = \sum_{q=0}^{Q} h^c \, \hat{x}_{t,k+12q}\,, \qquad (7)$$

where Q is the number of harmonics considered, and h is a positive value less than 1 that discounts the contribution of higher harmonics. We set the value of h to be 0.8 empirically. The result of summation is that the true note usually receives the largest amount of energy from its harmonic notes. Thus, the true note in frame t can be determined by choosing the note number associated with the largest value of the strength.

However, recognizing that a note usually lasts several frames, the decision could be made by including the information from neighbor frames. Specifically, we determine the note in frame t by choosing the note number associated with the largest value of the strength accumulated for adjacent frames, i.e.,

$$o_t = \arg\max_{1 \le k \le K} \sum_{w=-W}^{W} y_{t+w,k}\,. \qquad (8)$$

Further, the resulting note sequence is refined by taking into account the continuity between frames. This is done with median filtering, which replaces each note with the local median of notes of its neighboring $\pm W$ frames, to remove jitters between adjacent frames.

IV. Computation of Inter-Recording Similarities

After converting heart sounds into sequences of MIDI notes, we use bigram model [13] to capture the dynamic information in the note sequences. Here, the dynamic information is considered to be the major cause for why different heart beat categories sound different. The bigram model consists of a set of bigram probabilities and unigram probabilities. The bigram probabilities $\Pr(e_j|e_i)$, $1 \le i, j \le K$, accounts for the frequency of a certain note e_i followed by another note e_j, while the unigram probabilities $\Pr(e_i)$ accounts for the frequency of occurring a certain note e_i. It is assumed that each heart beat category has its own note sequence pattern that reflects in the frequency of occurrence of one or a pair of notes.

For N heart sound recordings to be clustered, we generate N bigram models $\lambda_1, \lambda_2, \ldots, \lambda_N$, and then compute the Cross Likelihood Ratio (CLR) [14] for every pair of recordings' note sequences $\mathbf{o}_r = \{o_{r,1}, o_{r,2}, \ldots, o_{r,T_r}\}$ and $\mathbf{o}_k = \{o_{k,1}, o_{k,2}, \ldots, o_{k,T_k}\}$:

$$CLR(\mathbf{o}_r, \mathbf{o}_k) = \log \frac{\Pr(\mathbf{o}_r \mid \lambda_k)}{\Pr(\mathbf{o}_r \mid \lambda_r)} + \log \frac{\Pr(\mathbf{o}_k \mid \lambda_r)}{\Pr(\mathbf{o}_k \mid \lambda_k)}, \qquad (9)$$

where

$$\Pr(\mathbf{o}_r \mid \lambda_k) = \prod_{t=1}^{Tr} \Pr(o_{r,t} \mid o_{r,t-1}, \lambda_k), \qquad (10)$$

and T_r and T_k are the lengths of sequences $\mathbf{o}_r$ and $\mathbf{o}_k$, respectively. It is assumed that the more the similar between $\mathbf{o}_r$ and $\mathbf{o}_k$, the larger the values of $\Pr(\mathbf{o}_r|\lambda_k)$, $\Pr(\mathbf{o}_k|\lambda_r)$, and $CLR(\mathbf{o}_r,\mathbf{o}_k)$.

V. Cluster Generation

After computing the inter-recording similarities, the next step is to assign the recordings deemed similar to each other to the same cluster. This is done by an agglomerative hierarchical clustering method [15], which consists of the following procedure:

```
begin: initialize M←N, and form clusters
   c_i←{o_i} , i = 1, 2, . . . , N
do
find the most similar pair of clusters, say
   c_i and c_j
merge c_i and c_j
M←M–1
until M = 1
end
```

The similarities between a pair of clusters, say c_i and c_j, can be derived from the inter-recording similarities, according to one of the following heuristic measures:

(i) *Complete linkage*

$$S(c_i, c_j) = \min_{\mathbf{o}_n \in c_i, \mathbf{o}_k \in c_j} CLR(\mathbf{o}_n, \mathbf{o}_k), \qquad (11)$$

(ii) *Single linkage*

$$S(c_i, c_j) = \max_{\mathbf{o}_n \in c_i, \mathbf{o}_k \in c_j} CLR(\mathbf{o}_n, \mathbf{o}_k), \qquad (12)$$

(iii) *Average linkage*

$$S(c_i, c_j) = \frac{1}{\#(\mathbf{o}_n \in c_i, \mathbf{o}_k \in c_j)} \sum_{\mathbf{o}_n \in c_i, \mathbf{o}_k \in c_j} CLR(\mathbf{o}_n, \mathbf{o}_k), \qquad (13)$$

where $\#(\mathbf{o}_n \in c_i, \mathbf{o}_k \in c_j)$ denotes the number of recording pairs involved in the summation.

VI. Estimation of the Number of Heart Beat Categories

Consider a collection of N sound recordings to be partitioned into M clusters. The optimal value of M must be an integer between 1 and N. Thus, if we produce a set of possible partitionings, in which the number of clusters ranges from 1 to N, the task of determining the optimal value of M would amount to selecting one of the N partitionings that achieves the level of within-cluster homogeneity as high as possible with the number of clusters as small as possible. To realize such a selection, we propose estimating the Rand Index for each of the possible partitionings, and then select the one that achieves the smallest value of Rand Index. This is based on the fact that the Rand index usually decreases with an increase in the value of M initially, and reaches the minimum at $M = P$. When $M > P$, the Rand index starts to increase as the value of M increases. This property motivates us to determine the population size by finding the value of M that achieves the minimal Rand Index.

Recalling the Rand Index in Eq. (3), the first term in the right side of the equation, $\sum_{m=1}^{M} n_{m*}^2$, can be computed based on the clustering result. The second term in the right side of the Eq. (3), $\sum_{p=1}^{P} n_{*p}^2$, is a constant irrelevant to clustering. However, the third term of Eq. (3), $\sum_{m=1}^{M}\sum_{p=1}^{P} n_{mp}^2$, requires that the true attribute of each sound recording be known in advance, which cannot be computed directly. To solve this problem, we represent $\sum_{m=1}^{M}\sum_{p=1}^{P} n_{mp}^2$ by

$$\begin{aligned}\sum_{m=1}^{M}\sum_{p=1}^{P} n_{mp}^2 &= \sum_{m=1}^{M}\sum_{p=1}^{P}\left[\sum_{i=1}^{N}\delta(h_i,m)\delta(a_i,p)\right]^2\\ &= \sum_{m=1}^{M}\sum_{p=1}^{P}\left[\sum_{i=1}^{N}\delta(h_i,m)\delta(a_i,p)\right]\left[\sum_{j=1}^{N}\delta(h_j,m)\delta(a_j,p)\right]\\ &= \sum_{m=1}^{M}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta(h_i,m)\delta(a_i,p)\delta(h_j,m)\delta(a_j,p)\\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\delta(h_i,h_j)\delta(a_i,a_j),\end{aligned} \tag{14}$$

where $\delta(\cdot)$ is a Kronecker Delta function, h_i is the index of cluster where the i-th recording is located, and a_i is the true attribute of the i-th recording. Note that h_i, $1 \le i \le N$, is an integer between 1 and M, if M clusters are generated. The term $\delta(a_i, a_j)$ in Eq. (14) is then approximated by the similarity between $\mathbf{o}_i$ and $\mathbf{o}_j$:

$$\delta(a_i,a_j) \approx \hat{\delta}(a_i,a_j) = \begin{cases} 1 & \text{, if } i = j \\ CLR(\mathbf{o}_i,\mathbf{o}_j)/CLR_{\max} & \text{, if } i \ne j, \text{and } CLR_{\max} > 0\text{,} \\ 0 & \text{, if } i \ne j, \text{and } CLR_{\max} = 0 \end{cases} \tag{15}$$

where $CLR_{\max}$ is the largest of the similarity measures $CLR(\mathbf{o}_i, \mathbf{o}_j)$, $\forall\ i \ne j$, which is used to making the value of $\hat{\delta}$ (a_i, a_j) between 0 and 1. In addition, recognizing the inevitable difference between $\delta(a_i, a_j)$ and $\hat{\delta}$ (a_i, a_j), we introduce a factor α in the estimation of the Rand index, which results in a criterion

$$\hat{R}(M) = \sum_{m=1}^{M} n_{m*}^2 - 2\alpha\sum_{i=1}^{N}\sum_{j=1}^{N}\delta(h_i,h_j)\hat{\delta}(a_i,a_j). \tag{16}$$

Hence, the optimal number of clusters can be determined by

$$M^* = \arg\min_{1\le M\le N} \hat{R}(M). \tag{17}$$

VII. EXPERIMENTS

The heart sound data used in this study were collected from several websites [16-18], including 91 sound recordings. The recordings could be divided into 9 heart beat categories manually, as shown in Table I. All the recordings were stored in 8000-Hz sampling rate and 16-bit quantization resolution.

TABLE I
HEART SOUND DATA USED IN THIS STUDY.

	Category	No. of Recordings
1	Normal heart sound	15
2	Aortic regurgitation	8
3	Aortic stenosis	9
4	Atrial septal defect	11
5	Mitral stenosis	10
6	Patent ductus arteriosus	12
7	Pericardial friction rub	7
8	Pulmonic stenosis	8
9	Ventricular septal defect	11

First, we examined the performance of our system under the condition that the number of heart beat categories is known *a priori*. The resulting performance could be served as an upper bound of the case that the number of heart beat categories is unknown. In this experiment, we set the number of generated clusters to be 9. Table II shows the experiment results, in which we compared the complete linkage, single linkage, and average linkage. It can be seen from Table II that the complete linkage is superior to the other linkages in this experiment. We obtained an average purity of 0.91 and clustering error rate of 11.2%.

TABLE II
CLUSTERING RESULTS UNDER THE CONDITION THAT THE NUMBER OF HEART BEAT CATEGORIES IS KNOWN A PRIOR.

Inter-cluster Similarity	Cluster Results	
	Average Purity	Clustering Error Rate
Complete Linkage	0.91	11.2
Single Linkage	0.71	31.4
Average Linkage	0.79	22.1

We then examined the system under the condition that the true number of heart beat categories is unknown and must be estimated. To simulate various numbers of heart beat categories involving in a data collection, we divided the database into three overlapping groups involving 3, 6, and 9 heart beat categories. The relationships between groups are shown in Fig. 1. We conducted clustering experiment for each groups separately, in order to examine if the optimal numbers of clusters determined using Eq. (17) could be close to 3, 6, and 9.

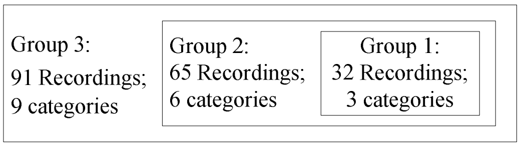

Fig. 1 Three groups of sound data involving different numbers of heart beat categories.

Table III shows the results of clustering each group. For the ease of comparison, we also list the clustering results obtained by setting the number of generated clusters to be the true number of heart beat categories, i.e., column "#Clusters = # True Heart Beat Categories". It can be seen from Table III that the optimal number of clusters determined automatically ("#Clusters = # Estimated Heart Beat Categories") is around the true number of heart beat categories involved in each group, despite a slight overestimate. We can also see from Table III that the purities and clustering error rate achieved with the estimated numbers of heart beat categories are close to those achieved with known numbers of heart beat categories. The results confirm the validity of the proposed heart sound clustering system.

TABLE III
RESULTS OF CLUSTERING THE 3 GROUPS

Groups	#Clusters = # True Heart Beat Categories Indicated within the Parenthesis	#Clusters = # Estimated Heart Beat Categories Indicated within the Parenthesis
Group 1: 3 categories	6.3 (3 Clusters)	5.0 (4 Clusters)
Group 2: 6 categories	10.1 (6 Clusters)	11.9 (8 Clusters)
Group 3: 9 categories	11.2 (9 Clusters)	19.5 (11 Clusters)

VIII. CONCLUSIONS

We have proposed an unsupervised framework for clustering unknown heart sound data such that each cluster contains sound recordings belonging to the same heart diseases or normal heart category. Such a framework is more flexible than the conventional supervised classification of heart sound data by the case when heart sound data belong to undefined categories or when there is no prior template data for building a heart sound classifier. Our experiments show that the resulting clusters based on our system are roughly consistent with the heart beat categories defined by human labeling, which indicates the feasibility of the unsupervised classification framework. Despite the potential, the methods proposed in this study can only be regarded as a preliminary investigation in realistic heart sound data clustering applications. In the future, more work is needed to study the effectiveness and efficiency of clustering a large scale of heart sound collection. In particular, it is necessary to deal with various heart sounds recorded from various environments and electronic stethoscopes.

REFERENCES

[1] Z. Sharif, M. S. Zainal, A. Z. Sha'ameri, and S. H. S. Salleh, "Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations," *in Proc. of TENCON*, Vol. 2 . pp. 130-134, 2000.

[2] J. Martínez-Alajarín, J. López-Candel, and R. R. Merino, "Classification and diagnosis of heart sounds and murmurs using artificial neural networks," *Lecture Notes in Computer Science*, Vol. 4527, pp. 303-312, 2007.

[3] Y. Chung, "Classification of continuous heart sound signals using the ergodic hidden Markov model," *in Proc. of 3rd Iberian conference on Pattern Recognition and Image Analysis.* Girona, Spain, Springer, 2007.

[4] Z. Dokur and T. Ölmez, "Heart sound classification using wavelet transform and incremental self-organizing map," *Digital Signal Processing*, Vol. 18, No. 6, pp. 951-959, 2008.

[5] W. C. Kao, C. C. Wei, and J. J. Liu, "Automatic heart sound analysis with short-time Fourier transform and support vector machines," *in Proc. of IEEE International Midwest Symposium on Circuits and Systems*, pp. 188-191, 2009.

[6] E. Khorasani, S. Doraisamy, A. Azman, and M. A. Murad, "An approach towards a heartbeat sound information retrieval system," *Computing in Cardiology*, Vol. 37, pp. 789-792, 2010.

[7] T. Topal, H. Polat, and I. Guler, "Software development for the analysis of heartbeat sounds with labview in diagnosis of cardiovascular disease," *Journal of Medical System*, Vol. 32, pp.409-21, 2008.

[8] E. Delgado-Trejos, A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Digital auscultation analysis for heart murmur detection," *Ann Biomed Eng*, Vol. 37, No. 2, pp. 337-53, 2009.

[9] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*. Vol. 66, pp. 846-850, 1971.

[10] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, Vol. 2, pp. 193–218, 1985.

[11] A. Solomonoff, A. Mielke, M. Schmidt, and H. Gish, "Clustering speakers by their voices," *in Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, 1998.

[12] M. Piszczalski and B. A. Galler, "Predicting musical pitch from component frequency ratios," *Journal of the Acoustical Society of America*, Vol. 66, No. 3, pp. 710–720, 1979.

[13] F. Jelinek, "Self-organized language modeling for speech recognition," *Readings in Speech Recognition*, Palo Alto, CA: Morgan Kaufmann, chap 8, pp. 450-506. 1990.

[14] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Communication*, Vol. 17, pp. 91-108, 1995.

[15] L. Kaufman, and P. J. Rousseuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, 1990.

[16] Advanced Physical Diagnosis in the University of Washington Department of Medicine: http://depts.washington.edu/physdx/heart/demo.html

[17] Dr R S MacWalter, University of Dundee, Ninewells Hospital & Medical School: http://www.dundee.ac.uk/medther/Cardiology/hsmur.html

[18] Bioscience: http://www.bioscience.org/atlases/heart/sound/sound.htm