NEURAL NETWORK AND CONVENTIONAL CLASSIFIERS TO DISTINGUISH BETWEEN FIRST AND SECOND HEART SOUNDS.

J. Edward Hebden* and J. N. Torry*

Summary

A technique to distinguish between the first and second heart sounds without the need for a reference ECG is described. The choice of features for presentation to classifiers is discussed and several types of classifier are introduced. Comparative results for each of the classification techniques are given for data sets obtained from both normal and pathological cases. A misclassification rate of 5.76% is obtained using a neural network classifier whereas conventional classifiers are shown to give a relatively poor performance.

Introduction

A wealth of useful diagnostic information is present in heart sound. This is clear from the continued use of the stethoscope by physicians during the physical examination. However, interpretation is often hampered by difficulties in distinguishing and classifying sounds because these skills take many years for an expert to acquire. An automated device to assist physicians with these tasks is feasible. The work of the heart sound group at Sussex has been directed towards the development of such a device.

Applications of advanced digital signal processing techniques to heart sound have been successful in extracting useful diagnostic features from the phonocardiogram waveform[1,2]. These techniques are typically only applied to short sections of the sound which must first be isolated from the rest of the cardiac cycle. The results of this work have been to develop a technique to perform this segmentation.

In normal subjects only the first and second heart sounds (S1 and S2) are usually appreciated. In pathological cases, abnormal sounds such as murmurs, and third or fourth heart sounds may be detected in addition to S1 and S2. S1 and S2 can be isolated from abnormal sounds by bandpass filtering[3]. A problem arises, however in distinguishing between S1 and S2 since they are in the same frequency range and are morphologically very similar. The ECG is usually recorded simultaneously with the heart sound in order that this distinction can be made[4,5]. The QRS complex of the ECG arrives shortly before the first heart sound and since this can be detected with relative certainty the first heart sound can be distinguished from the second. This paper describes an approach to the problem of automatic segmentation that does not rely on recording the ECG reference. If the need to record the ECG can be eliminated the procedure to record heart sounds is simplified. In addition, electrical isolation requirements are eliminated since there is no longer any electrical contact with the patient.

The approach taken is to isolate S1 and S2 from other sounds, and then to differentiate between them by extracting features from the sounds and presenting these to a classifier. The choice of classifier is critical for the success of the technique. Classifiers based on statistical techniques and a neural network trained by backpropagation have been implemented. This paper compares the relative performance of the classifiers for different data sets.

The structure of the paper is as follows: the next section describes data acquisition; this is followed with a

*From the Graduate Division of Biomedical Engineering, University of Sussex.

description of pre-processing to identify the beginning and end of each S1 or S2. The choice of features to extract from each sound is then justified on the basis of trends observed in heart sound recordings. Both neural network and conventional classifiers are described and results are presented to compare their performances for the available data sets.

Data Collection.

The recordings have been made with a system developed to allow reliable simultaneous recording of the heart sound and ECG. Specialised contact sensors are interfaced to a PC via a proprietary interface card. Appropriate amplification and analogue filtering stages confine the signal to the frequency range 20 - 750Hz. This removes large scale low frequency signals and prevents aliasing. A 12 bit analogue-to-digital conversion is then performed at a sampling frequency of 4096Hz.

Automatic Detection of S1 and S2

A method of pre-processing has been applied to each of the recordings to distinguish S1 and S2 from other sounds. The aim of the technique is to establish the start and end point of each S1 and S2 in a recording so that features can then be extracted. *Figure 1* shows a block diagram of this process. S1 and S2 are generally attributed to vibrations of heart tissue initiated by deceleration of the blood mass following valve closure[6]. Murmurs are thought to arise from turbulent blood flow through stenosed or insufficient valves. Murmurs typically contain higher frequency components than S1 and S2, and can therefore be removed by filtering. Firstly, an FFT is applied to a single cycle of the recording and a dominant first peak, characteristic of S1 and S2 is obtained. Higher frequency peaks due to any murmurs may also be present. The location of the first peak is used to set the cut-off frequencies of a band-pass filter which removes high frequency murmurs and any low frequency sounds from the signal. The resulting waveform is converted to an envelope using a Hilbert transform. Each sample is then multiplied by the two previous samples to accentuate any peaks in the recording, this makes choosing a suitable threshold simpler. A

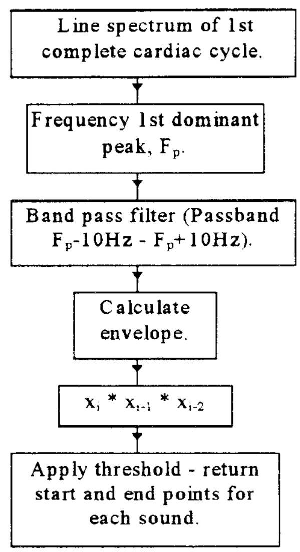
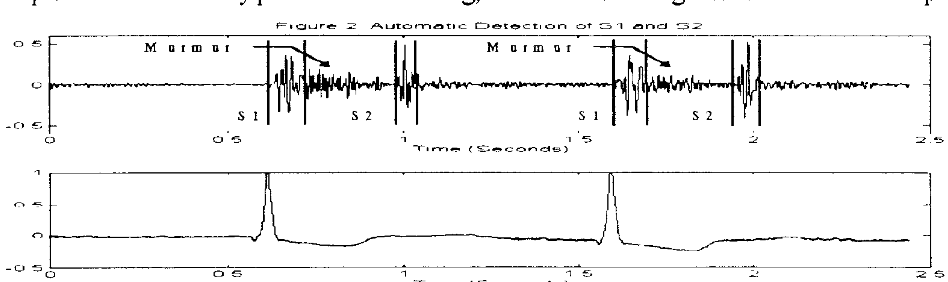


Figure 1: **Automatic detection of S1 and S2 prior to feature extraction.**

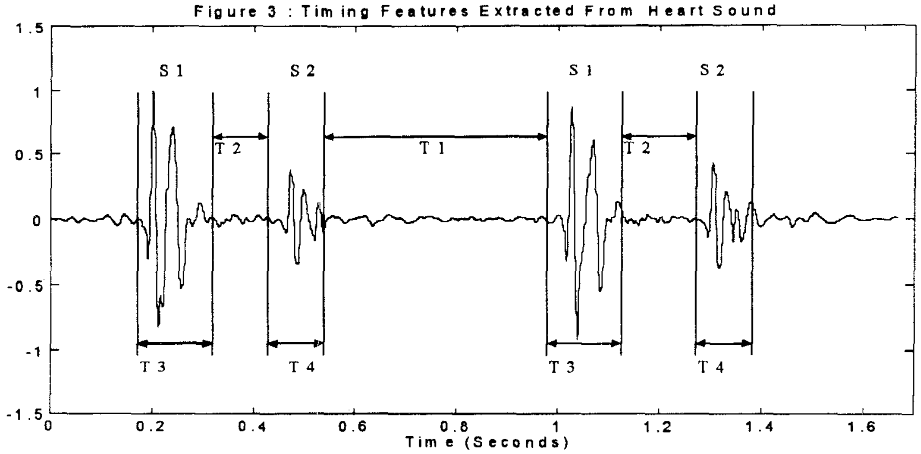


threshold is chosen manually. The lowest threshold for all normal sounds to be detected without a false positive detection is selected by a process of trial and error. *Figure 2* shows a few cycles of heart sound and ECG. The vertical lines show the start and end of each S1 and S2, estimated using the automatic detection algorithm.

Feature Extraction.

The choice of features is based on observing trends in the timing of S1 and S2 in recordings from "normal" subjects. A few seconds from such a recording is shown in *Figure 3*.

Figure 3 : Timing Features Extracted From Heart Sound



This recording is typical of a healthy heart sound. Several generalisations can be made about the waveform:

1. The time from the end of S1 to the beginning of S2 is shorter than the time from the end of the S2 to the beginning of the S1 ($T_1 > T_2$).
2. The duration of the S1 is greater than the duration of the S2 ($T_3 > T_4$).
3. The amplitude of the S1 is greater than the amplitude of the S2.

A fourth difference between S1 and S2 that is not obvious from *Figure 3*, is that the energy distribution with time is not always symmetrical for S2 - this is due to splitting. Splitting arises when the closure of valves is asynchronous,

Figure 4a: Split S2: top - waveform, bottom - envelope

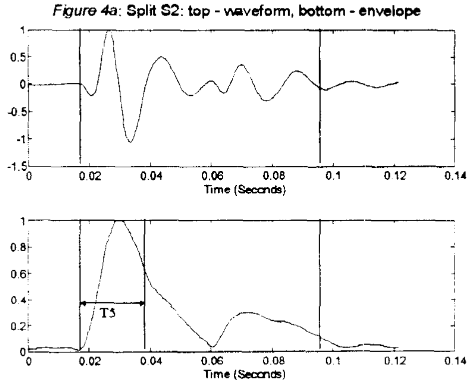
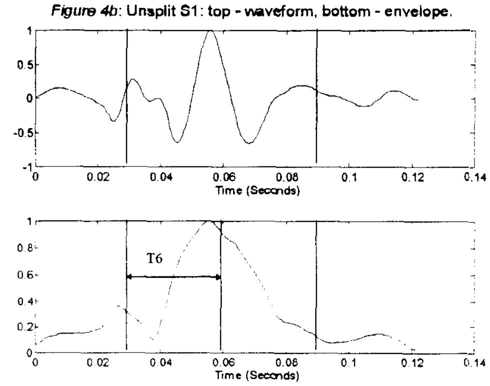


Figure 4b: Unsplit S1: top - waveform, bottom - envelope.



the sounds resulting from the closure of each valve superimpose to produce a "split" S2. The sounds are rarely of the same amplitude and therefore the resulting energy envelope has a skew. Splitting of S2 is usually observed in normal subjects as well as in pathological cases, however it is less frequently seen in S1. *Figure 4a* shows a split S2 together with its skewed energy distribution, *Figure 4b* shows an unsplit S1 and its symmetric energy distribution. A measure of the skewness of the envelope waveform can be obtained by calculating the cumulative sum for each sample in the envelope. Each sample is added to the sum of the previous samples. The number of the sample at which the 50th percentile occurs, divided by the total number of samples gives a measure of the skewness.

These trends are not observed in every case - when a subject is stressed or diseased, or if the recording position is changed one or more of the relationships may change. Nevertheless, using experience gained from observing these trends, qualitative tests showed that it is usually possible for a human observer to distinguish S1 from S2 without reference to the ECG. It has been widely reported that neural networks are capable of making this type of generalisation, it was for this reason that it was decided to apply neural network techniques to this task.

Four features are therefore extracted, they are summarised in *Table 1* below for the third sound in *Figure 3*. For features 2, 3 and 4 ratios are taken of the feature for the current sound to the same feature for the next. This is because it is known in advance that consecutive sounds must be different.

Table 1: Summary of features.	
(1) $F1 = \frac{T1}{T2}$	(2) $F2 = \frac{T3}{T4}$
(3) $F3 = \frac{\frac{\sum_{i=1}^{M3} x_i^2}{T3}}{\frac{\sum_{i=1}^{M4} x_i^2}{T4}}$	(4) $F4 = \frac{\frac{T6}{T3}}{\frac{T5}{T4}}$

x_i are the data points, M3 and M4 the number of points in the intervals T3 and T4 respectively, T5 and T6 are the times to the 50th percentile of the cumulative sum in intervals T3 and T4 respectively.

Classification.

The aim of the classification procedure has been to develop a rule whereby any new observation, represented by the feature vector, **x**, can be classified into one of the classes S1 or S2. Since it is known in advance which observations in the database of recordings correspond to S1 and S2 (from the reference ECG) supervised learning methods can be applied. Initially a rule based approach was implemented to compare features for successive sounds. The performance of this technique was found to be poor compared to the other classifiers. It may be possible to improve the performance with a closer inspection of the data set and construction of more detailed rules. However statistical and neural network approaches usually generalise better to unseen data than rule based methods and therefore this approach was not pursued further.

In all the tests described below 50% of the data set was chosen at random to train the classifier and the remaining "unseen" 50% used to test the classifier. A method of cross validation was applied to estimate the error rate for each classifier. Many repeat runs were carried out in which different random selections of training and test sets were chosen. An average error rate was then calculated from all the runs. For the classifier to best learn the problem the training data set should be as large as possible. This implies using the "leave one out" method in which all but one of the examples is used for training, the remaining example being used for testing. All possible combinations are tried and the average error rate calculated. However training times can become prohibitively long when training some classifiers with large data sets in this way and so a less rigorous approach, using 50% of the data, was taken. Below is a brief description of each of the classifiers used in the study, a more detailed description of the methods is given by Mitchie et. al.[7]:

1. **Linear Discriminant**: the aim is to divide the sample space into two by a hyperplane. Classification for an unseen case then depends on which side of the hyperplane the example falls. The hyperplane is chosen to separate the known classes as well as possible. This is done using a method of maximum likelihood.
2. **Nearest Neighbour**: The Euclidean distance, e, from the test observation to all the training observations is calculated. The class of the test observation is taken to be the same as the training observation to which it is nearest, i.e. for which e is smallest. A slightly more sophisticated version was also implemented in which **the k nearest neighbours** to the test observation were found and the classification made by a voting system. The value of k was chosen for the smallest misclassification rate.
3. **Naive Bayes:** The assumption is made that the features are independent. The probability that a given observation, **x**, is a member of a given class is calculated for each feature. The calculation is based on an estimate of the probability density function (pdf) for each feature considered separately. The probabilities from each of the features are then multiplied together and the class with the greater resulting value is chosen. The pdfs are estimated from the training data, either by assuming a normal distribution and calculating the mean and variance, or by dividing the distribution into bands and calculating the relative frequency in each band.
4. **Non Parametric Density Estimation:** In this approach the features are not assumed to be independent and no assumption is made about the pdfs. The pdf's for each class, $f_{S1}(x_i)$ and $f_{S2}(x_i)$, are estimated from the training data using a kernel function and suitable smoothing parameter[8]. The class of the test observation is taken to be that for which the calculated pdf is the greatest.
5. **Neural Network**: The widely used feedforward network trained by backpropagation is implemented in this study[9]. A method using momentum and an adaptive learning rate was used to speed up learning. Both two and three layer networks with various numbers of hidden neurons were trained. In order to minimise training times the networks were tested as they were trained. Every 100 epochs, the test data were presented to the network and the percentage misclassified was calculated. When this figure stopped decreasing the network was considered "trained" and training stopped. A few runs with very long training times were performed to check that further training did not significantly reduce the misclassification rate.

Comparison of Performance of Classifiers.

Two sets of recordings were made for this work: both the heart sound and ECG being recorded in each case to enable the evaluation of the results of the method by a "gold standard". Recordings from 10 subjects with normal heart sound, and from 12 subjects with a variety of valvular disorders were made. Altogether a total of 2600 cardiac cycles were used in the study. The results are summarised in *Table 2* below.

Table 2: Misclassification rates for each classifier applied to each data set.

	Normals (%)	Abnormals (%)	All recordings (%)
Linear Discriminant.	0.04 +/- 0.07	13.86 +/- 8.05	6.58 +/- 3.60
Nearest Neighbour.	0.02 +/- 0.08	17.74 +/- 4.65	8.63 +/- 2.44
K - Nearest Neighbour.	0.05 +/- 0.11	13.94 +/-2.52	6.62 +/- 3.20
Naive Bayes.	0.01 +/- 0.03	13.65 +/- 3.24	7.13 +/- 1.38
Non Parametric Density Estimation.	0.20 +/- 0.47	10.63 +/- 3.37	7.60 +/- 2.94
Neural Network.	0.65 +/- 0.83	9.69 +/- 2.48	5.76 +/- 1.32

For each classification technique half the data were chosen at random to train the classifier and the remaining half used to test the classifier. Three data sets were considered separately: normal subjects, abnormal subjects and both abnormal and normal subjects together. For each data set between 50 and 100 runs with different random selections of training and test data were performed with each classification technique. The table gives the mean and standard deviation of the minimum percentage of misclassifications obtained for each run.

Conclusions.

A technique to distinguish between the normal heart sounds, S1 and S2, without the need for a reference ECG has been described. Various approaches to the classification stage of the problem have been described. The performance of each technique when applied to data sets obtained from normal subjects, abnormal subjects and both abnormal and normal cases together have been given. The neural network approach gives lower misclassification rates than the "conventional" approaches for all three data sets, with a misclassification rate of 5.76% for the combined data set of abnormal and normal cases. This result seems to confirm the findings of other researchers that neural networks perform better than conventional classifiers for some classification tasks. Since processing is off line it is possible that this misclassification rate could be reduced by considering the classification of several sounds in a recording.

Caution should be exercised when assessing the result because the data set used is relatively small (containing data from 22 subjects only). Other heart disorders, not included in the data set, may produce variations in the features that are not observed in the recordings we have studied. It could then prove difficult for classifiers to generalise to these new observations. Obviously, further testing with a larger data set would be desirable. One extension to the work is to use unsupervised neural networks, or clustering techniques, to organise the data into groups, this could lead to subsets within the classes being revealed and therefore simplification of classifiers. It is possible that this simplification might reduce the misclassification rate Further work is continuing in this direction, as well as testing the classifiers with a larger data set.

Acknowledgements.

We are grateful to Dr. C. Davidson, Dr. A. Haider and their patients at the Royal Sussex County Hospital in Brighton for their assistance in making the recordings for this study. We also wish to thank the Trafford Centre for Medical Research at the University of Sussex for their support of this work.

References.

1 Joo, T. H., McClellan, J. H. et. al. "Pole - zero modelling and classification of phonocardiograms". IEEE Transactions on Biomedical Engineering. Vol. BME 30. No. 2. February 1983, 110-117.

2 Durand, L. G., De Guise, J., et. al. "Evaluation of FFT - based and modern parametric methods for the spectral analysis of bioprosthetic valve sounds". IEEE transactions on Biomedical Engineering. Vol. BME 33. No. 6, June 1986, 572-78.

3 Iwata, A., Ishii, N., Suzumura, N., "Algorithm for detecting the first and second heart sounds by spectral tracking". Med. & Biol. Eng. & Comput. 1980, 18, 19-26.

4 Baranek, H. L., Lee, H. C., et. al. "Automatic detection of sounds and murmurs in patients with Ionescu-Shiley aortic bioprostheses". Med. & Biol. Eng. & Comput. 1989, 27, 449-455.

5 Haghighi - Mood, A. Torry, J. N. "A sub - band energy tracking algorithm for heart sound segmentation". Proceedings of IEEE Computers in Cardiology, Vienna. 1995, 501-504.

6 Luisada, A. A., Portaluppi, F. "The main heart sounds as vibrations of the cardiohemic system: old controversy and new facts". American Journal of Cardiology. Vol. 52. 1 November 1983, 1133-1136.

7 Mitchie, D., Spiegelhalter, D. J., Taylor, C. C. "Machine learning, neural and statistical classification". Ellis Horwood Ltd. 1994

8 Duin, R. P. W. "On the choice of smoothing parameters for Parzen estimators of probability density functions". IEEE transactions on computers. November, 1976, 1175-1179.

9 Rumelhart, D. E., Hinton, G. E. "Learning internal representations by error propagation". In: "Parallel Distributed Processing", (Rumelhart, D. E. & McClelland, J. L., eds) Volume 1, 318-362. MIT Press Cambridge MA.