

PHONOCARDIOGRAM SEGMENTATION BY USING AN HYBRID RBF-HMM MODEL

Carlos S. Lima, Manuel J. Cardoso

Department of Industrial Electronics of University of Minho, Campus de Azurém, Guimarães, Portugal
carlos.lima@dei.uminho.pt

Abstract

This paper is concerned to the segmentation of heart sounds by using Radial-Basis Functions for acoustical modelling, combined with a Hidden Markov Model for heart sounds sequence modelling. The idea behind the use of RBF's is to take advantage of the local approximations using exponentially decaying localized nonlinearities achieved by the Gaussian function, which increases the clustering power relatively to MLP's. This neural model can be advantageous over the global approximations to nonlinear input-output mappings provided by Multilayer Perceptrons (MLP's), especially when non-stationary processes need to be accurately modelled.

The above described RBF's properties combined with the non-stationary statistical properties of Hidden Markov Models can help in the detection of the T-wave which is fundamental for the detection of the second heart sound.

The feature vectors are based on a MFCC based representation obtained from a spectral normalisation procedure, which showed better performance than the MFCC representation alone, in an Isolated Speech Recognition framework. Experimental results were evaluated on data collected from five different subjects, using CardioLab system and a Dash family patient monitor. The ECG leads I, II and III and an electronic stethoscope signal were sampled at 977 samples per second.

Keywords: Hidden Markov Models, Radial-Basis functions, Phonocardiogram segmentation, spectral normalisation.

1. Introduction

The phonocardiogram (PCG) is a sound signal related to the contractile activity of the cardiohemic system. The general state of the heart in terms of contractility and rhythm can be provided by heart sounds characteristics. Cardiovascular diseases and defects cause changes or additional sounds and murmurs that can be useful for diagnosis purposes. A normal cardiac cycle contains two major sounds: the first heart sound S1 and the second heart sound S2. S1 occurs at the onset of ventricular

contraction and corresponds in timing to the QRS complex hence it can be easily identifiable if the ECG is available which is frequently the case. S2 follows the systolic pause and is caused by the closure of the semilunar valves. The interval between S1 and S2 as well as the S2 sound are both very important concerned to the diagnosis of several pathologies such as valvular stenosis and insufficiency. It is well known, for example, that S1 is loud and delayed in mitral stenosis, right bundle-branch block causes wide splitting of S2 and left bundle-branch block results in reversed splitting of S2 [1,2,3]. Murmurs are noise-like events, which can appear in the systolic segment, in the interval between S1 and S2 and in the diastolic segment representing obviously different pathologies. Although they are all noise-like events their features aid in distinguishing between different causes. For example, aortic stenosis causes a diamond-shaped midsystolic murmur whereas mitral stenosis causes a decrescendo-crescendo type diastolic-presystolic murmur. Automatic diagnosis of these and many others cardiovascular defects or diseases require robust techniques for segmenting the phonocardiogram. Especially the detection of S2 sound is hard to obtain in spite of it appears slightly after the end of the T-wave, however, as the T-wave is often a low amplitude and smooth wave and sometimes not recorded at all, thus the T-wave is not a reliable indicator to use for the identification of S2.

Traditional techniques for S2 detection are mainly based on the notch in the aortic pressure wave, which can be obtained by using catheter tip tensors [4,5], which is an invasive procedure. Fortunately, the notch is transmitted through the arterial system and may be observed in the carotid pulse recorded at the neck. The dicrotic notch in the carotid pulse signal will bear a delay with respect to the corresponding notch in the aortic pressure signal but has the advantage of being accessible in a noninvasive manner. This delay is sometimes taken into consideration in the detection of S2 [6]. Signal processing techniques for the detection of the dicrotic notch and segmentation of the phonocardiogram include, among others, least-squares estimate of the second derivative of the carotid pulse [6], averaging techniques [7] and more recently the use of heart sound envelopeogram, which reports a 93% success

rate. However, implementing this algorithm is prone to error and it is sensitive to changes in pre-processing and setup parameters, which strongly compromises its robustness.

Recently new approaches based on pattern recognition have been applied in solving difficult problems concerned to classification purposes, such as automatic speech recognition, cardiac diagnosis and segmentation of medical images, among others. These algorithms rely heavily on parametric signal models, which parameters are learned from examples. The most common approaches of this class are the Neural Networks (NN) approach and the Hidden Markov Model (HMM) approach.

This paper reports the use of RBF's, which is a type of NN appropriated for statistical modeling, combined with HMM's, as a robust model regarding to the segmentation of the phonocardiogram into its component segments: the S1 sound, the systole period, the S2 sound and the diastole period.

2. PCG Features Extraction

The features extraction method considers noise robustness, is based on the power spectral density components and consists in dividing the power spectral density inside each sub-band by the total short-time power. The power in each sub-band is obtained by summing the power spectrum components inside the sub-band. All the sub-bands have the same number of spectral components and no one is shared by different sub-bands, thus avoiding increases of statistical dependence between sub-bands (feature components). This kind of normalisation seems to be also adequate for dealing with additive distortions since the numerator and denominator of the features are both increased, though by different values, however this fact contributes for stabilising the feature values, which means increasing the robustness.

To best understand this reasoning, consider S_i denoting the power in sub-band i and S denoting the short time signal power of the considered segment. Similarly, let N_i and N denote the power of the interfering noise in sub-band i and the short time noise power, respectively. So, the i^{th} component of the observation vector for the clean signal is given by

$$c_i = \frac{S_i}{S} \quad (1)$$

Similarly if the signal appears noisy the next equation holds

$$c_{in} = \frac{S_i + N_i}{S + N} \quad (2)$$

where the index n stands for noisy signal. Equations (1) and (2) are computed in the same way without concerning

to the noise existence, so they can be viewed as the same equation. The denominators of equations (1) and (2) represent respectively the power of the signal segment in clean and noisy conditions and can be both computed by summing all the components of the power spectrum density.

If the interfering noise has white noise characteristics the environment will shift the clean vector by a noise dependent vector $C_i(N)$, which can be computed by subtracting equation (1) from equation (2).

If the noise is stationary then its short time power equals its long time power. Note that this does not occur for the phonocardiogram due to its non-stationary property, but as an approximation we will consider that the short time phonocardiogram signal power equals the long time phonocardiogram signal power. Under this constraint, S and N can be related by the signal to noise ratio (SNR). Therefore the next expression holds

$$S + N = S \left(1 + \frac{1}{10^{\frac{SNR}{10}}} \right) \quad (3)$$

Let l , the number of components in each sub-band and L the FFT length. Then N and N_i , considering flat noise spectrum, are related by the quotient l/L . By using these considerations, the calculation of the shift vector imposed by the environment to the observed vector component i is noise dependent and is accomplished by subtracting equation (1) from equation (2)

$$\begin{aligned} C_i(N) &= \frac{S_i + N_i}{S \left(1 + \frac{1}{10^{\frac{SNR}{10}}} \right)} - \frac{S_i}{S} = \frac{S_i - kS_i}{kS} + \frac{N_i}{kS} \\ &= \left(\frac{S_i}{S} - \frac{N_i}{N} \right) \frac{1-k}{k} \end{aligned} \quad (4)$$

where k is given by

$$k = 1 + \frac{1}{10^{\frac{SNR}{10}}} \quad (5)$$

and in terms of mean the next expression holds

$$N_i = \frac{l}{L} N \quad (6)$$

Equation (4) shows that if the clean signal has a flat power spectrum density, the means of $C_i(N)$ become null as S_i/S equals l/L . This is particularly true for example in processing unvoiced regions of the speech signal. For the

present case of phonocardiogram processing we expect that this feature can help in the segmentation since the segment between S1 and S2 contains only noise or noise-like signals (murmurs) respectively for a normal and abnormal subjects. Therefore, this normalisation process becomes optimal in the sense that the environment does not affect the means of the phonocardiogram features, while the variances are strongly reduced by the intrinsic mechanism of energy normalization, which consists of the mathematical division of the power in each sub-band by the short-time power.

The normalized spectrum obtained in this manner is then applied to a Mel-Spaced filter bank. Mel-Spaced filter banks provide a simple method for extracting spectral characteristics from an acoustic signal, and are largely used in the field of speech processing. This method involves creating a set of triangular filter banks across the spectrum. The filter banks are equally spaced along the Mel-scale as defined by equation (7)

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (7)$$

Equal spacing on the Mel-scale provides exponential spacing on the normal frequency axis. This exponential spacing means that there are numerous small banks at lower frequencies and sparse, large banks at higher frequencies. Since most of the phonocardiogram energy is in the lower frequency ranges, using a Mel-scale matches the frequency spectrum of the heart sounds.

Each triangular filter is multiplied by the normalised spectrum obtained from equation (1) and summed for each triangular filter. This constitutes the feature vectors to be used as HMM observations.

Since the average duration of S1 is about 0.16 seconds, the HMM segments the phonocardiogram in frames of 0.15 seconds with a 0.015 seconds frame overlapping. Each signal segment has 146 samples, which are converted into 256 spectral coefficients by using a conventional FFT algorithm. The signal is then divided in 16 sub-bands each one with 16 spectral coefficients. Then equation (1) is applied to perform spectral normalization, and the spectrum obtained in this way is applied to the Mel-scale filter banks. Since spectral dynamics is very important concerned acoustic signal modelling delta coefficients are computed and inserted in the observation vectors.

It is well known that the frequency spectrum of S1 contains peaks in the 10 to 50 Hz range and the 50 to 140 Hz range, while the frequency spectrum of S2 contains peaks in the 10 to 80 Hz range, the 80 to 200 Hz range and the 220 to 400 Hz range. Hence, this study limits the spectral feature extraction between the frequencies of 10 Hz and 420 Hz.

3. Heart Sound Model

The cardiac sound model is a composition of various RBF's, each one modeling each different sound unit. The different sound units are the silence and the major cardiac sounds S1 and S2. Each RBF provides a probability (likelihood) of generating each sound unit, which is assumed to be the emission probability associated with each HMM transition.

3.1. Heart Sound Markov Model

A Hidden Markov Model (HMM) is a probabilistic state machine where hidden (unobservable) states output observations. In the case of Discrete Density Hidden Markov Models (DDHMM's) the observations are discrete symbols from a finite alphabet. This kind of HMM can model the phonocardiogram as it traverses a specific labeled region such as S1, systolic, S2 or diastolic segments.

The segmentation of the phonocardiogram can be heuristically accomplished by the discrete HMM which four state (one state for each event) structure is shown in figure 1.

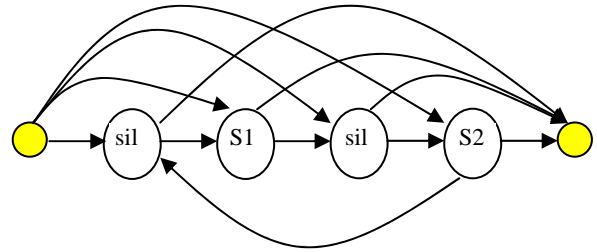


Figure 1. Heart sound Markov Model

This HMM does not take into consideration the S3 and S4 heart sounds since these sounds are difficult to hear and record, thus they are most likely not noticeable in the records.

3.2. Acoustic Model

The acoustic model is RBF network based. The method of radial-basis functions is a technique for interpolation in a high dimensional space and provides an alternative to learning in the neural networks framework. The RBF network consists of an only hidden layer of high enough dimension which provides a non-linear transformation from the input space. The output layer provides a linear transformation from the hidden-unit space to the output space. The activation function $f(x)$ called RBF responds to a field of view around a fixed location c such that the function is largest for $x=c$ and decreases (approaches zero) as x becomes more distant of c . Usually the considered distance is the Euclidean distance and the activation function is the Gaussian function. Figure 2 shows the Probabilistic neural network architecture.

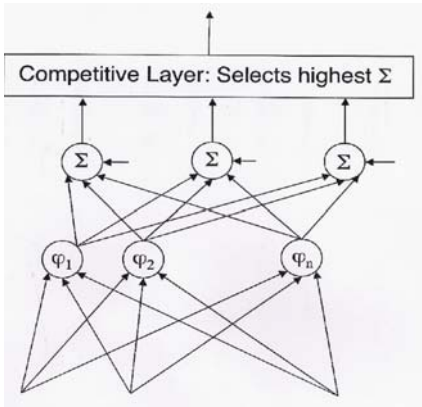


Figure 2. PNN used for acoustic modelling

When using radial-basis function neurons, a category of patterns can be regarded as a Gaussian distribution of points in pattern space. The RBF neuron fires when its input is sufficiently close to activate the Gaussian. For this reason the RBF's are usually known as NN's for statistical applications.

4. Experimental Results

Experimental results were evaluated by using five records from different subjects. The results were computed on the basis of frame error rate where each frame of the labelled signal was compared to the output signal. The system error rate was computed by dividing the number of mismatched frames by the total number of frames in the system.

Three from the existing five records were used for training purposes after labelling.

Another evaluation was done on the basis of model inaccuracies. The difference between the centre of the heart sound label and the centre of the learned heart sound was computed. The maximum value allowed for this delta was 40 milliseconds. The model error rate was computed as the number of mismatched S1 or S2 labels divided by the total number of sound labels in the system. The testing set is composed by the five records, so it includes the training set. Table 1 shows the results separately for the training set and for the two remaining records, which represents really the test set.

Table 1 – Error rates.

Testing set	Frame error rate	Model error rate
Training records	0.055±0.024	0.029±0.012
Remaining records	0.11±0.082	0.098±0.063

6. Conclusions

The main objective of the work described in this paper was to develop a robust segmentation technique for segmenting the phonocardiogram into its main components. In a variety of experiments RBF's showed some advantages over MLP's such as faster training, better retention of generalization, more accuracy, less sensitivity to bad training data, to name a few. The performance obtained was slightly better than the one reported in [8]. However, the performance of our system can be increased by increasing the training set.

References

- [1] Rangayyan, R. M. and Lehner, R. J. (1988). Phonocardiogram signal processing: A review. *CRC Critical Reviews in Biomedical Engineering*, 153 (3): 211-236.
- [2] Travel, M. E. (1978). *Clinical Phonocardiography and External Pulse Recording*. Year Book Medical, Chicago, IL, 3rd edition.
- [3] Luisada, A. A. and Portaluppi F. (1982). *The Herat Sounds – New Facts and Their Clinical Implications*. Praeger, New York.
- [4] Shaver, J. A., Salerni, R., and Reddy, P. S. (1985). Normal and abnormal heart sounds in cardiac diagnosis, Part I: Systolic sounds. *Current problems in Cardiology*, 10(3): 1-68.
- [5] Reddy, P. S., Salerni, R., and Shaver, J. A., (1985). Normal and abnormal heart sounds in cardiac diagnosis, Part II: Diastolic sounds. *Current problems in Cardiology*, 10(4): 1-55.
- [6] Lehner, R. J. and Rangayyan, (1987). A three-channel microcomputer system for segmentation and characterization of the phonocardiogram. *IEEE Transactions on Biomedical Engineering*, 34:485-489.
- [7] Durand, L. G., de Guise, J., Cloutier, G., Guardo, R. and Brais, M. (1986). Evaluation of FFT-based and modern parametric methods for the spectral analysis of bioprosthetic valve sounds. *IEEE Transactions on Biomedical Engineering*, 33(6):572-578.
- [8] Lang, H., Lukkarinen, S. and Hartimo, I. (1997). Heart Sound Segmentation Algorithm Based on Heart Sound Envenlogram. *Computers in Cardiology*; 105-8.