

PHONOCARDIOGRAM SEGMENTATION BY USING HIDDEN MARKOV MODELS

Carlos S. Lima , Manuel J. Cardoso

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

ABSTRACT

This paper is concerned to the segmentation of heart sounds by using state of art Hidden Markov Models technology. Concerning to several heart pathologies the analysis of the intervals between the first and second heart sounds is of utmost importance. Such intervals are silent for a normal subject and the presence of murmurs indicate certain cardiovascular defects and diseases. While the first heart sound can easily be detected if the ECG is available, the second heart sound is much more difficult to be detected given the low amplitude and smoothness of the T-wave. In the scope of this segmentation difficulty the well known non-stationary statistical properties of Hidden Markov Models concerned to temporal signal segmentation capabilities can be adequate to deal with this kind of segmentation problems.

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, 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 could be useful in their diagnosis. 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 sensors [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 envelope, 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. From a theoretical point of view HMM's are more adequate for modeling time event sequences, especially when the events appear in the same sequence, however this potential advantage is not always proved in practical applications.

This paper reports the use of HMM's, as a robust technique for segmenting 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 Si/S equals 1/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. HIDDEN MARKOV MODELS

A Hidden Markov Model (HMM) is a probabilistic state machine where hidden (unobservable) states output observations. In the case of Continuous Density Hidden Markov Models (CDHMM's) the observations are modeled by mixture densities, usually Gaussian, such that the probability density function for each HMM state/transition is given by

$$f(\mathbf{x}) = \sum_{c=1}^C p_c G(\mathbf{x}, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \quad (8)$$

where C is the number of components in the mixture, p is the mixture coefficient (weight), $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are respectively the mean vector and covariance matrix of the c^{th} mixture component and G stands for Gaussian function. This kind of HMM can model the phonocardiogram as it traverses a specific labeled region such as S1, systolic, S2 or diastolic segments.

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

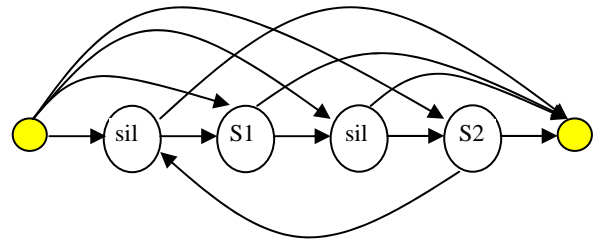


Figure 1. Heart sound Markov Model

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.

In order to model accurately the continuous transitions between sound and silence a CDHMM which structure is shown in figure 2 is embedded in the model shown in figure 1. This model structure is used in speech recognition systems where the transition of words or phonemes is modeled by an HMM similar to the one shown in figure 1, while phonemes are modeled by a structure similar to the one shown in figure 2.

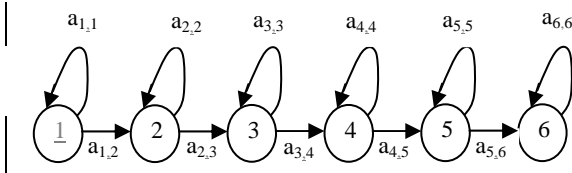


Figure 2. HMM topology for acoustic modelling.

Concerned to the more adequate number of HMM states to model a given signal a rule does not exist. However it is very common to use a three state model for pauses and/or silences in speech recognition applications since in these regions the non-stationary nature of the signal is usually not very strong. Therefore in our experiments the silences were modelled by a three state HMM.

Since our acoustic model has only transitions for the next state and not for example for the next of the next, a model which number of states is larger than the number of observations does not makes sense once the model must begin on the first state and ends on the last state for each observation sequence. For example the S1 duration is about 160 ms and the frame step size is 15 ms, therefore a 10 state HMM can segment the S1 sound in just 10 segments, one segment for each observation. Usually some parts of the signal are quasi-stationary and a small number of states is usually considered. In our case we used 6 states and 3 mixture components in each state transition.

As S1 sound is longer than S2 we used 4 states in the modelling of S2 and also 3 mixture components in each state transition.

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 | 0.053±0.021 | 0.031±0.011 |
| Remaining records | 0.12±0.084 | 0.096±0.061 |

5. 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. 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. Additionally ambient noise can be reduced by using state of the art techniques appropriate for HMM modelling [9].

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.
- [9] C. Lima, L. B. Almeida and J. L. Monteiro, "Continuous Environmental Adaptation of a Speech Recogniser in Telephone Line Conditions," 7th International Conference on Spoken Language Processing (ICSLP'2002), pp 1401-1404, 2002.