Automatic Segmentation of the Second Cardiac Sound by Using Wavelets and Hidden Markov Models

C. S. Lima and D. Barbosa

Abstract— **This paper is concerned with the segmentation of the second heart sound (S2) of the phonocardiogram (PCG), in its two acoustic events, aortic (A2) and pulmonary (P2) components. The aortic valve (A2) usually closes before the pulmonary valve (P2) and the delay between these two events is known as "split" and is typically less than 30 miliseconds. S2 splitting, reverse splitting or reverse occurrence of components A2 and P2 are the most important aspects regarding cardiac diagnosis carried out by the analysis of S2 cardiac sound. An automatic technique, based on discrete wavelet transform and hidden Markov models, is proposed in this paper to segment S2, to estimate de order of occurrence of A2 and P2 and finally to estimate the delay between these two components (split). A discrete density hidden Markov model (DDHMM) is used for phonocardiogram segmentation while embedded continuous density hidden Markov models are used for acoustic models, which allows segmenting S2. 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.**

I. INTRODUCTION

THE phonocardiogram (PCG) is a sound signal, which carries important information about the general state of contractile activity of the cardiohemic system.. 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. S2 follows the systolic pause and is caused by the closure of the semilunar valves. The importance of S2 regarding diagnosis purposes has been recognized for a long time, and its significance is considered of utmost importance, by cardiologists, to auscultation of the heart [1].

This paper will concentrate mainly on the analysis of the second heart sound (S2) and its two major components A2 and P2. The main purposes are estimating the order of occurrence of A2 and P2 as well as the time delay between them. This delay known as "split" occurs from the fact that the aortic and pulmonary valves do not close simultaneously. Normally the aortic valves close before the pulmonary valves and exaggerated splitting of the S2 sound

may occur in right ventricular outflow obstruction, such as pulmonary stenosis (PS), right bundle branch block (RBBB) and atrial and ventricular septal defect. Reverse splitting of sound S2 is due to a delay in the aortic component A2, which causes a reverse sequence of the closure sounds, with P2 preceding A2. The main causes of reverse spitting are left bundle branch block (LBBB) and premature closure of pulmonary valves. The wide "split" has duration of about 50 miliseconds compared to the normal "split" with the value of ≤ 30 ms [2]. The measurement of the S2 "split", lower or higher than 30 ms and the order of occurrence of A2 and P2 leads to a discrimination between normal and pathological cases.

Most biomedical signals tend not to be stationary, typically having highly complex time frequency characteristics. Frequently they consist of brief, high frequency components closely spaced in time, accompanied by long lasting, low frequency components closely spaced in frequency. Usual Fourier methods are adequate for picking out frequencies from a signal consisting of many frequencies, but they are utterly incapable of dealing properly with a signal that is changing over time. Because of their compatibility with non-stationary random processes the wavelet transform (WT) is a powerful tool for analysing biomedical signals. The good time frequency localization is the most important advantage that wavelets have over some other methods. Furthermore, WT has demonstrated the ability to analyse the heart sound more accurately than other techniques as the spectrogram (STFT) or Wigner distribution (WD) [3] in some pathological cases. One of the key issues regarding signal parameterization in the scope of wavelet transform is the selection of the scales which are the least affected by noise maintaining, however the most useful signal information. This topic is of particular interest concerning phonocardiogram analysis since murmurs, which are noise-like events can appear in both diastolic and systolic segments. These murmurs, which are very important concerned to the diagnosis of several pathologies such as valvular stenosis and insufficiency, if not attenuated, seriously difficult the automatic detection of the cardiac sounds specially when only the phonocardiographic signal is available. In the scope of this paper, three consecutive dyadic wavelet scales were selected. The choice was made on the basis of the frequency content of S1, S2, A2 and P2. It is well known that the sound S2 has higher frequency content than S1, and A2 has also higher frequency content than P2. Therefore, it is possible to select an appropriate level of focus, which depends on the sampling frequency, or

in other words an appropriate scale for the which these frequency content differences are enhanced.

Phonocardiogram analysis requires frequently robust detection of S1 and S2 sounds. Traditional techniques for this purpose are mainly based on the simultaneous recording of other signals such as the ECG (Electrocardiogram) and CP (Carotid pulse). S1 occurs at the onset of ventricular contraction and corresponds in timing to the QRS complex. Therefore, it can be easily identifiable if the ECG is available. Detection of S2 is more difficult since the T-wave is not a reliable indicator of its identification. In fact T-wave is often a low amplitude and smooth wave and sometimes not recorded at all. A much more reliable indicator of the identification of the S2 sound is the notch in the aortic pressure wave, which is transmitted through the arterial system and may be observed in the carotid pulse recorded at the neck. 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 [4], averaging techniques [5] and more recently the use of heart sound envenlogram, 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 with 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. Hidden Markov models (HMM) are a class of these parametric models that can be used to model phenomena in which deterministic observed symbols are arranged in temporal series. An HMM consists of a set of states connected by transitions between them. At each time instant a transition takes place and the model generates one observation. In this manner a non-stationary signal can be modeled by a quasi-stationary signal, since the first is sliced in stationary segments. It has been shown that HMM's have the capability of breaking the phonocardiogram in its main components S1 and S2 [6] [7] especially under situations of weak murmurs. The presence of strong murmurs claims for robust parameterization techniques. This paper shows that also the constituent waves of the main sounds can be accurately modeled by HMM's if an appropriate grammar is selected. This study can also be easily extended to the S1 sound, however this is out of the scope of this paper. In the scope of this paper three different scales of the wavelet transform are simultaneously observed by a continuous density hidden Markov model.

II. WAVELETS

The wavelet transform (WT) is a signal representation in a scale-time space, where each scale represents a focus level of the signal and therefore can be seen as a result of a band-pass filtering.

Given a time-varying signal x(t), WTs are a set of coefficients that are inner products of the signal with a family of "wavelets" obtained from a standard function known as "mother wavelet". In Continuous Wavelet Transform (CTW) the wavelet corresponding to scale "s" and time location "τ" is given by

$$\psi_{\tau,s} = \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-\tau}{s}\right) \tag{1}$$

where $\psi(t)$ is the mother wavelet, which can be viewed as a band-pass function. The term $\sqrt{|s|}$ ensures energy preservation. In the CWT the time-scale parameters vary continuously

$$\Psi_x^{\psi}(\tau,s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right) dt \tag{2}$$

where the asterisk stands for complex conjugate. Equation (2) shows that the WT is the convolution between the signal and the wavelet function at scale "s". Therefore the shape of the mother wavelet seems to be important in order to emphasize some signal characteristics, however this topic is not explored in the ambit of the present work.

For implementation purposes both "s" and "τ" must be discretized. The most usual way to sample the time-scale plane is on a so-called "dyadic" grid, which means that sampled points in the time-scale plane are separated by a power of two. This procedure leads to an increase in computational efficiency for both WT and Inverse Wavelet Transform (IWT). Under this constraint the Discrete Wavelet Transform (DCT) is defined as

$$\psi_{j,k}(t) = s_0^{-j/2} \psi\left(s_0^{-j}t - k\tau_0\right) \tag{3}$$

which means that DWT coefficients are sampled from CWT coefficients. ASS "dyadic" scale is used and therefore $s_0=2$ and $\tau_0=1$, yielding $s=2^j$ and $\tau=k2^j$ where j and k are integers.

As the scale represents the level of focus from the which the signal is viewed, which is related to the frequency range involved, then digital filter banks are appropriated to break the signal in different scales (bands). If the progression in the scale is "dyadic" the signal can be sequentially half-band high-pass and low-pass filtered.

The output of the high-pass filter represents the detail of the signal. The output of the low-pass filter represents the approximation of the signal, for each decomposition level, and will be decomposed in its detail and approximation components at the next decomposition level, and the process proceeds iteratively in a scheme known as wavelet decomposition tree, which is shown in figure 1. After the filtering half of the samples can be eliminated according to

335

the Nyquist's rule, since the signal now has only half of the frequency.
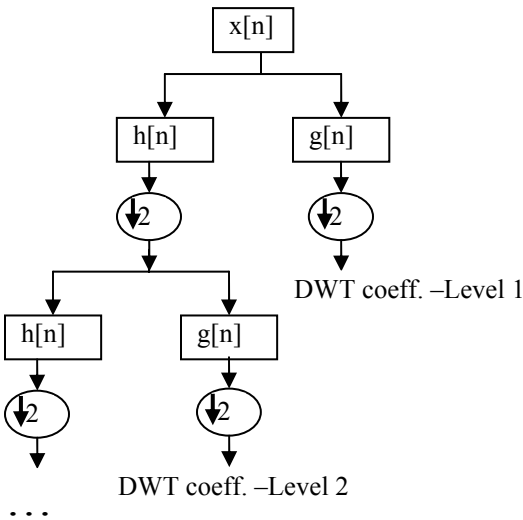

Figure 1- Wavelet decomposition tree

This very practical filtering algorithm yields as Fast Wavelet Transform (FWT) and is known in the signal processing community as two-channel subband coder [8].

One important property of the DWT is the relationship between the impulse responses of the high-pass (g[n]) and low-pass (h[n]) filters, which are not independent of each other and they are related by

$$g[L-1-n] = (-1)^n h[n] \qquad (4)$$

where L is the filter length in number of points. Since the two filters are odd index alternated reversed versions of each other they are known as Quadrature Mirror Filters (QMF). Perfect reconstruction requires, in principle, ideal half-band filtering. Although it is not possible to realize ideal filters, under certain conditions it is possible to find filters that provide perfect reconstruction. The most famous ones were developed by Ingrid Daubechies and they are known as Daubechies' wavelets. In the ambit of this work only Daubechies' wavelets with 2 vanishing moments (db-4) were used.

III. WAVELET ANALYSIS OF THE PHONOCARDIOGRAM

The major difficulty associated with the phonocardiogram segmentation is the similarity among its main components. For example it is well known that S1 and S2 contain very closed frequency components [9], however S2 have higher frequency content than S1. Another example of sounds containing very closed frequency components, which must be distinguished is the aortic and pulmonary components of S2 sound. The multiresolution analysis based on the DWT can enhance each one of these small differences if the signal is viewed at the most appropriate scale. Figure 2 shows the result of the application of the DWT one cycle of a normal PCG. From the figure we can observe that d1 level (frequency ranges of 250-500Hz) emphasize the high frequency content of S2 sound when compared with S1. D2 and d3 levels show clearly the A2 component, while d4

represents clearly the main components of S1, named M1 and T1. The features used in the scope of this work are simultaneous observations of d1, d3 and d4 scales, therefore the observation sequence generated after the parameter extraction is of the form $\mathbf{O}=(\mathbf{o}_1, \mathbf{o}_2, \dots \mathbf{o}_T)$ where T is the signal length in number of samples and each observation $\mathbf{o}_t$ is a three-dimensional vector, i. e., the wavelet scales have the same time resolution as the original signal.
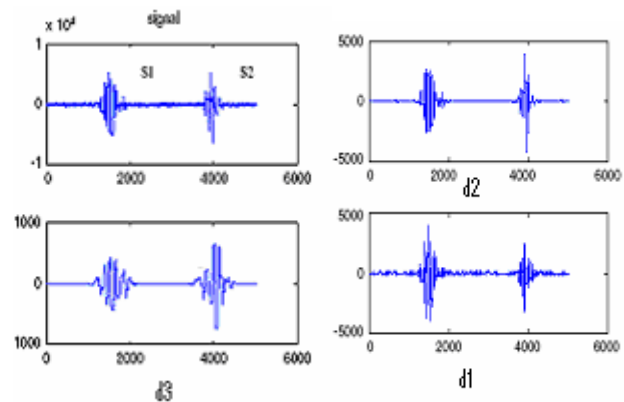

Figure 3. Wavelet decomposition of the normal cardiac sounds. Upper panel on the left shows the original signal.

IV. HMM SEGMENTATION OF THE PCG

The phonocardiogram can be seen as a sequence of elementary waves and silences containing at least five different segments; M1, T1, A2, P2 and silences. Each one of them can be modeled by a different HMM. In the scope of this paper we used left to right (or Bakis) HMM's with different number of states, since this is the most used topology in the field of speech recognition, and the phonocardiographic signal is also a sound signal with auditory properties similar to speech signals. Each HMM models one different event and the concatenation of them models the whole PCG signal. The concatenation of these HMM's follows certain rules dependent on the sequence of events allowed. These rules define a grammar with four main symbols and an associated language model as shown in figure 4.
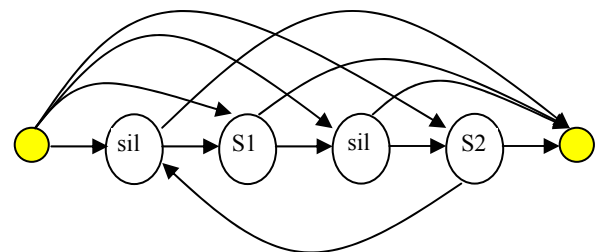

Figure 4. Heart sound Markov Model

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.

336

The acoustic HMM's are Continuous Density Hidden Markov Models (CDHMM's) and the continuous observations are frequently modeled by a Gaussian mixture. However, by observing the histograms for every state of every HMM we observed that most of them appear to be well fitted to a single Gaussian, so we used single Gaussian probability density functions to model the continuous observations.

The PCG morphologies are learned by training the HMM's. The training algorithm was the standard Baum-Welch method, also called forward-backward algorithm, which is a particular case of the expectation maximization method. Each waveform model was trained on a set of approximately 40 patterns.

The beat segmentation procedure consists on matching the HMM models to the ECG waveform patterns. This is typically performed by the Viterbi decoding algorithm, which relates each observation to an HMM state following the maximum likelihood criteria with respect to the beat model structure. Additionally the most likely state sequence is available, which allows to estimate time duration of the PCG components as the split.

V. EXPERIMENTAL RESULTS

Experimental results were evaluated by using eight records from different subjects. The segmentation 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. The obtained performance was high, about $99,1\pm0.02\%$ over a set with about 700 cardiac pulses. Automatic Diagnosis was performed by estimating the order of occurrence of A2 and P2 as well as the split. Concerning to S1 no diagnostic actions were taken. Murmurs were not classified in the scope of this work.

Table1- Results of PCG segmentation

Pathologies	Detected	Non-detected	Wrong detected
A2 after P2	34	2	3
Split > 45 ms	8	1	2

Three from the existing eight records were used for training purposes after labelling. The testing set is composed by the remaining five records, so does not include the training set. No records present severe murmurs. Table 1 shows the results. Long splits were calculated by using a 2 state HMM for split modelling. As the sampling frequency is about 1 kHz, if the HMM remains in these 2 states more than 45 times the split is excessively long. This information is provided by the Viterbi algorithm. A2 and P2 are modelled by a 3 state HMM and the inverse occurrence of them is directly computed by the grammar of the application.

VI. CONCLUSION

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. HMM's proved that they can perform well not only in segmenting the PCG in its main components but also in segmenting the components of the main components allowing automatic diagnosis related with abnormal order of appearance of the components of the main PCG sounds.. Additionally signal durations can be robustly estimated which is of utmost importance in order to diagnose split related diseases. The discrete wavelet transform can be efficiently used to improve signal discrimination specially when frequencies with minimal differences need to be emphasized. The main drawback of this approach is that the performance of the method degrades significantly in severe murmurs, specially in aortic and mitral regurgitation.

REFERENCES

[1] Leatham, A. (1987). Auscultation and Phonocardiography: a personal view of the past 40 years. Heart J. 57 (B2).
[2] Leung, T. S., White, P. R., Cook, J., Collis, W. B., Brown, E., Salmon, A.P., (1998). Analyse of the second heart sound for diagnosis of paediatric heart disease. IEE Proc. Sci. Meas . Technol. 145 (6) 285-290.
[3] Obaidat, M. S. (1993). Phonocardiogram signal Analysis: techniques and performance comparison. J. Med. Eng. Technol. 17 (6) 221-227.
[4] 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.
[5] 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.
[6] Lima, C. S. and Cardoso, M. J. (2007). Phonocardiogram Segmentation by Using Hidden Markov Models. The Fifth IASTED International Conference in Biomedical Engineering, BioMED 2007, Innsbruck Austria.
[7] Lima, C. S. and Cardoso, M. J. (2007). Phonocardiogram Segmentation by Using an Hybrid HMM-RBF Model. The Fifth IASTED International Conference in Biomedical Engineering, BioMED 2007, Innsbruck Austria.
[8] Rangayyan, R. M. and Lehner, R. J. (1988). Phonocardiogram signal processing: A review. CRC Critical Reviews in Biomedical Engineering, 153 (3): 211-236.
[9] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, M. Maldonado, R Schmidt and J Habetha (2006). Wavelet Transform and Symplicity Based Heart Murmurs Segmentation.: Computers in Cardiology; 33:173-176.