A NEW ALGORITHM FOR DETECTION OF S1 AND S2 HEART SOUNDS

D. Kumar, P. Carvalho, M. Antunes†, P. Gil, J. Henriques, L. Eugénio†

Centre for Informatics and Systems, University of Coimbra, Coimbra, Portugal
†Centre of Cardio-thoracic Surgery of the University Hospital of Coimbra, Coimbra, Portugal
E-mail: {dinesh, carvalho, pgil, jh }@dei.uc.pt, antunes.cct.huc@sapo.pt

ABSTRACT

This paper presents a new algorithm for segmentation and classification of S1 and S2 heart sounds without ECG reference. The proposed approach is composed of three main stages. In the first stage the fundamental heart sound lobes are identified using a fast wavelet transform and the Shannon energy. Next, these lobes are validated and classified into S1 and S2 classes based on Mel-frequency coefficients and on a non supervised neural network. Finally, regular heart cycles are identified in a post-processing stage by a heart rhythm criterion. This approach was tested using sound samples collected from prosthetic valve implanted patients. Results are comparable with ECG based approaches.

1. INTRODUCTION

Many heart disorders can be effectively diagnosed using auscultation techniques. In potentially deadly heart diseases, such as natural and prosthetic heart valve dysfunction or even in heart failure, heart sound auscultation is one of the most reliable and successful tools for early diagnosis. To develop automatic heart disorder diagnosis tools based on phonocardiogram analysis, it is required to first segment the heart sound into clinically meaningful segments or lobes, such as the S1 and the S2 sound components associated with closing valves during systole and diastole. Diagnostic features (e.g. timbre) can subsequently be extracted from these lobes.

Heart sound segmentation algorithms can broadly be classified into two classes: algorithms which are based on ECG reference and those not requiring this signal. The first type algorithms use the QRS complex and the T-wave to locate the precise instances of S1 and S2 components, respectively. Variants can be found for which low quality ECG signals without clear or absent T-waves are used [1]. Despite being highly robust, this class of approaches exhibit additional hardware requirements and impose some constraints to patient's comfort. Algorithms which are not based on ECG synchronization may be further classified into supervised and unsupervised methods. Several methods can be found in the literature where supervised classification techniques (see e.g. [2] and unsupervised methods (see e.g. [3]) are applied. Unsupervised methods use some empirical thresholds and take into account several assumptions regarding the loudness of S1 and S2. These methods have already achieved promising results for heart sounds produced by native valves, but tend to lack invariance to recording location. Nevertheless, it is well known that assigned features to heart sound components are dependent on patients, site of auscultation, artifacts and surgery techniques, in the case of prosthetic heart valves. Furthermore, for heart sounds produced by artificial heart valves noticeably different amplitude and spectral characteristics are observed. Consequently, methods based on fixed empirical thresholds or supervised learning approaches tend to provide ambiguous sound lobe identification. One solution to this problem could be fine tuning the algorithm for each patient and in addition could be less time consuming.

In this paper an unsupervised S1 and S2 heart sound segmentation algorithm is presented, which does not require ECG synchronization. The proposed approach is composed of three main stages. First the fundamental heart sound lobes are identified using Shannon energy computed from the low frequencies of the signal identified by means of the fast wavelet transform coefficients. Next, these sound lobes are validated using physiologically inspired criteria and further characterized based on jitter, in order to discriminate among heart sounds and sound lobes mixed with noise (for instance, swallowing or speech). The identified heart sound lobes are then classified into S1 and S2 classes using a self-organizing map classifier based on the Mel-frequency cepstral coefficients. Finally, in the last stage a post processing based on a heart rate criterion is applied.

2. METHODS

2.1. Segmentation by Fast Wavelet Transform

The goal of this stage is to clearly identify the boundaries of all sound lobes presented in phonocardiograms. This is

This project was partially financed by the IST FP6 project IST-2002-507816 supported by the European Union and POSI (Programa Operacional da Sociedade de Informação) of the Portuguese government supported by the European Union.

achieved by first low pass filtering the signal followed by the computation of its envelope. Boundary allocation of sound lobes is carried out by analyzing zerocrossing of the normalized envelope. In order to avoid empirical tuning of a filter bank, the fast wavelet transform (FWT) is applied to remove high frequency components from the sound samples. Here the db6 from Daubechies wavelet family is selected because of its orthogonal property, accuracy and computational inexpensiveness.

The FWT is performed until the 5th level. For further processing only the approximation coefficients are considered, which represent the transient part of the low frequency heart sound obtained using this procedure. This decomposition depth was experimentally found. To extract the signal envelop from the approximation coefficients the Shannon energy (1) is computed according to,

$$E_S = -\frac{1}{N}\sum_{i=1}^{N}(\phi^5_{S_{norm}})^2 log(\phi^5_{S_{norm}})^2, \qquad (1)$$

where $\phi^5_{S_{norm}}$ stands for the normalized 5th level FWT approximation coefficients ϕ^5_S of the heart sound signal S and N is the number of samples in the selected window.

This technique emphasizes the medium intensity signal and attenuates the effect of low intensity signals much more than the high intensity signals. Hence, it is better than normal energy or the absolute value in finding differences between the low intensity and high intensity sound [3].

In order to compute (1), the normalization with respect to the maximum absolute value of ϕ^5_S is taken. Shannon energy is computed using a centered sliding window of 20 ms with 50% overlap. Using the signal envelopes provided by the Shannon energy, sound lobe boundaries are identified from zerocrossing of the normalized Shannon energy, i.e.

$$E_{S_{norm}} = E_S - < E_S >, \qquad (2)$$

where '<>' represents the average operator.

2.2. Heart Sound Segments Validation

As it can be observed from figure 1, many irrelevant sound lobes are identified with the described procedure. For instance, very low/large duration segments as well as segments containing noise may not be conveniently filtered out using the FWT and the Shannon energy approach. These irrelevant sound segments are removed at this stage using four criteria based on physiologically inspired properties. In order to validate the sound lobes, there are three prime features of heart sound that are taken into consideration: the sound lobe duration, interval between two consecutive sound lobes and their loudness (root mean square). Let n_i^{start} be the start sample index and n_i^{stop} the stop sample index of a given segment, then they are computed according to,

$$dt^i = T_s(n_i^{stop} - n_i^{start}),\ T_i^{int} = T_s(n_{i+1}^{start} - n_i^{stop}), \quad (3)$$

$$RMS_i = \sqrt{\frac{\sum_{j=n_i^{start}}^{n_i^{stop}} S(j)^2}{n_i^{stop} - n_i^{start}}}, \quad i = 1, 2.....segments, \qquad (4)$$

where T_s is the sampling period. For each sound lobe the following validation steps are considered: I) The duration of S1 and S2 sounds is not more than 250 ms and not less than 30 ms in normal population or even in cardiac patients. Hence, segments for which durations are outside this interval may be considered as noisy sound segments and, therefore, discarded for further processing.

II) Interval T^{int} between two consecutive segments is considered to identify splitting in diastoles. It is seen that two sound segments which are separated by an interval less than 50 ms may belong to the same second heart sound. From the viewpoint of prosthetic valve dysfunction analysis the aortic sound is more important than the pulmonary sound, having the later always lower loudness than the former. This criterion is considered so as to discard pulmonary sound lobes.

III) Sometimes it is observed that splitting in the second heart sound may not be clearly captured by the Shannon energy due to low loudness of high frequency sounds. However, in this circumstances pronounced local minima in the positive Shannon energy are clearly found. If this local minima is less than 25% of the maximum value of the Shannon energy of the sound segment under analysis, then splitting in the second heart sound is considered to occur at that position. In these conditions, only the subsegment that shows the highest loudness is kept for further processing.

IV) During heart sound acquisition, high frequency long duration sounds may be mixed with heart sounds. In order to identify such sound segments corrupted with periodic disturbances jitter is computed [4]. It represents the perturbation in stochastic and temporarily long sustained sounds (e.g. swallowing and speech) when valves do not close normally or due to other external disturbances.

Let $T_p(k)$, $k = 1, 2, ...N_T$ be the time separation between the kth and the $(k+2)th$ local maxima of the autocorrelation of the heart sound signal then jitter in heart sound segments is computed according to,

$$Jitter = \frac{\sum_{k=2}^{N_T-1} 2T_p(k) - T_p(k-1) - T_p(k+1)}{\sum_{k=2}^{N_T-1} T_p(k)}, \quad (5)$$

Autocorrelation is computed with a 50 ms sliding and centered window, which is sufficient to get three consecutive peaks in noisy segments. If for a given window of the sound lobe jitter is detected, the lobe is marked as noisy and discarded for further processing. Some results using this procedure are shown in figure 1.

2.3. Classification of heart sounds S1 and S2

All sound lobes provided by the previous stage (segmentation) are classified here into S1 and S2 classes. This task is ac-

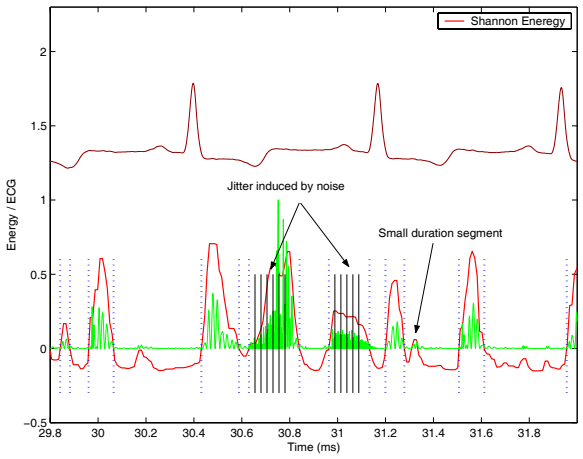

Fig. 1. Heart sound segmentation.

complished using an unsupervised self organizing map neural network (SOM) based on most discriminative feature among sounds, i.e. mel-frequency coefficients (MFCC), which are extracted for each sound lobe. In order to calculate MFCC, heart sound segments are first passed through a mel-based filter bank, converted to the mel-scale and then submitted to cosine reduction. This filter bank is constructed using 13 linearly-spaced filters (133.33 Hz between center frequencies), followed by 27 log-spaced filters. Each of the 12 channels (CH1-CH12) produces a set of MFCC. The dimension of this feature space is reduced by summing all coefficients produced by each channel. Thus, each channel will output exactly one feature for each sound lobe. The features produced by channel CH0 are not applied, since its large variance would dominate the organization of the SOM. The algorithm of classification using SOM is described in the following steps.

Step 1: The goal is to remove outliers by panning off the very low and very high loudness segments from the training set. The loudness of all the segments are transformed by Box-Cox transformation, which transforms a non-normal distribution into a normal one. Let λ be the value that maximizes the Log-Likelihood Function (LLF), then the transformation is given by,

$$RMS(\lambda) = \begin{cases} \frac{RMS^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ log(RMS) & \text{if } \lambda = 1 \end{cases} \quad (6)$$

For training proposes, only sound segments which verify $\mu - \sigma < RMS(\lambda) < \mu + \sigma$ are taken, where μ and σ are the mean and standard deviation of RMS.

Step 2: The SOM is trained using the MFCC feature vector of the training set. The size of SOM is determined by the dimension of training data set which is a great advantage over feed forward artificial neural networks.

Step 3: The map is clustered into two clusters using a k-means approach. Distinction between S1 and the S2 classes is performed based on physiological observations. For an extensive range of heart rates, it is seen that the time interval between S2 and S1 is higher than the time interval between S1 and S2. To implement this step, the interval (see (3)) for the two classes is first averaged and the class that shows the higher average is considered as S2.

Step 4: The outliers removed in step 1 are classified according to the identified classifier in step 2 and 3. This step is performed in order to find heart sound lobes in the removed outliers.

2.4. Post-processing by Heart Rate

This stage is devoted to identifying sound lobe sequences that comply with a set of physiologically inspired criteria. Namely, the heart rate is nearly constant during the recording of heart sound, since the patient is usually at rest; the diastolic interval is always greater than the systolic interval; the range of heart rate (in adults) at rest is usually between 60 and 100 beats per minute Thus, the duration of one heart cycle typically lies between 600 ms and 1000 ms.

Step 1: The heart rate is computed from correctly identified heart beats in the classified training set. In the process of computing heart rate, first, all the $S1 \rightarrow S2 \rightarrow S1$ and $S2 \rightarrow S1 \rightarrow S2$ sequences in the classified sounds are found, as depicted in figure (2), and the duration (T_B) of these heart cycles are computed by,

$$T_B = \frac{dt^i}{2} + T_i^{int} + dt^i + T_{i+1}^{int} + \frac{dt^{i+2}}{2}, \quad (7)$$

All the heart cycles T_B that exhibit a duration between 600 ms and 1000 ms are taken for further processing.

Step 2: From the detected beats, systolic and diastolic intervals are calculated. Let dt_{S1}^i and dt_{S2}^i be the durations of S1 and S2 of two consecutive sound lobes, then the systolic interval T_{sys} and the diastolic interval T_{dia} are calculated from,

$$T_{sys} = \frac{dt_{S1}^i}{2} + T_i^{int} + \frac{dt_{S2}^{i+1}}{2}, \; T_{dia} = \frac{dt_{S2}^{i+1}}{2} + T_i^{int} + \frac{dt_{S1}^{i+2}}{2}, \quad (8)$$

The median of all achieved T_{sys} as well as of T_{dia} are taken for further processing.

Step 3: Duration of two contiguous segments is calculated according to (9).

$$T_{temp}^i = \frac{dt^i}{2} + T_i^{int} + \frac{dt^{i+1}}{2}, \quad (9)$$

To accommodate some natural variation in heart rhythm and arrhythmic cases, a tolerance in heart rate is assumed. In the current algorithm implementation this tolerance is set as $\pm 15\%$ (tol=0.15).

For each contiguous pair of segments T_{temp}^i is compared with respect to T_{sys} and T_{dia}. For correct classification of S1

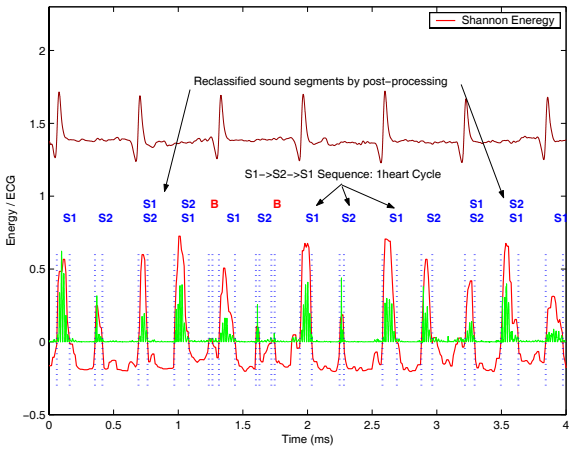

Fig. 2. Result of post-processing. *B* represents bad segments.

and S2, if condition (10) holds, then i^{th} and $(i+1)^{th}$ segments are S1 and S2 respectively. If condition (11) holds, then the i^{th} and the $(i+1)^{th}$ are S2 and S1 respectively.

$$(1-tol)T_{sys} < T^{i}_{temp} < (1+tol)T_{sys} \qquad (10)$$

$$(1-tol)T_{dia} < T^{i}_{temp} < (1+tol)T_{dia} \qquad (11)$$

If none of these conditions are verified, it indicates that the segment is irrelevant or a given segment is missing. In these circumstances, the next segment is picked up and step 3 is repeated. Some results using these approaches are presented in figure 2.

3. RESULTS AND DISCUSSION

Heart sounds have been recorded from different patients with prosthetic valve implants (both Mechanical and Bioprosthetic) one month after surgery using an electronic stethoscope from Meditron. All sound samples were digitized with 16 bit resolution and 44.1 kHz sampling rate. Next, they were preprocessed by a 4th order high pass butterworth filter with 40 Hz cutoff frequency to remove very low frequencies produced by slow movements, such as muscles or chest movements. The test database was composed of 49 sound samples taken from different patients, 9 with bioprosthetic and 40 mechanical valve implants in mitral and aortic position.
In table 1, some results obtained for 10 heart sounds samples are shown. It includes also the worst case result obtained for the test set (see sample C20S2 which is an arrhythmic sample). The overall sensitivity and specificity achieved for the database was 94.77% (with standard deviation of 4.97%) and 96.16% (with standard deviation 2.20%) respectively. In the experiments 365 segments were detected as false negative and 116 as false positive, out of 6614 segments. These are significant results, comparable to those achieved using an ECG reference for segmentation and classification [1].

Mechanical Valve Sound Samples(position)	S1 (detected/ present)	S2 (detected/ present)	False Negative (FN) / Sensitivity	False Positive (FP) / Specificity
C17S0 (Aortic)	180/182	182/182	2 / 99.45%	0 / 100%
C20S2 (Aortic)	49/69	38/38	20/81.31%	3/96.67%
C48S1 (Aortic)	56/56	56/56	0/100%	0/100%
C61S0 (Aortic)	67/68	66/66	1/99.25%	0/100%
C2S8 (Mitral)	37/39	36/39	4/94.67%	2/97.23%
Bioposthetic valve				
C51S1 (Aortic)	38/39	39/39	1 /98.73%	1 /98.73%
C36S2 (Aortic)	91/95	95/95	4 /97.94	2 /98.99
C42S4 (Mitral)	81/91	82/88	13 /92.61%	7 /95.88%
C8S1 (Aortic)	90/100	92/100	18/91%	5/97.33%
C14S7 (Aortic)	101/113	105/113	20/91.87%	3/98.69%

Table 1. Results form some sound samples.

4. CONCLUSIONS

In this paper, an algorithm for S1 and S2 heart sound identification, without an ECG reference, has been presented. The segmentation is accomplished by Shannon energy of the fast wavelet transformed heart sound. The S1 and the S2 sounds were classified by SOM and identified by the regular pattern of systolic and diastolic intervals. For reclassification of missclassified segments physiological based criteria were implemented. One of the main features of the proposed approach is its invariance to recording location and independence on particular sound characteristics. The achieved results are comparable to those obtained using ECG based approaches.

5. REFERENCES

[1] P. Carvalho, P. Gil, J. Henriques, M. Antunes and L. Eugénio, "Low complexity algorithm for heart sound segmentation using the variance fractal dimension," in *Proc. of the Int. Sym. on Intelligent Signal Processing*, 2005.

[2] J. E. Hebden and I. N. Torry, "Neural network and conventional classifiers to distinguish between first and second heart sound," *Artificial Intelligence Methods for Biomedical Data Processing, IEE Coll.*, vol. 3, pp. 1–6, 1996.

[3] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Proc. of IEEE Computers in Cardiology*, pp. 105 – 108 , 1997.

[4] D. L. Thomson and R. Chengalvarayan, "Use of periodicity and jitter as speech recognition features," in *Proc. of the ICASSP*, pp. 21 – 24, 1998.