HEART SOUND SEGMENTATION BASED ON MEL-SCALED WAVELET TRANSFORM

Karim Courtemanche[1], Véronique Millette[2], Natalie Baddour[2]
[1]*Faculty of Science, McGill University*
[2]*Department of Mechanical Engineering, University of Ottawa*

ABSTRACT

The identification and segmentation of heart beats in a phonocardiogram signal is of great interest, with applications ranging from diagnosis to use as a timing source. This paper proposes a new algorithm for the identification of the first and second (S1 and S2) heart sounds and for segmentation of the signal. The proposed algorithm is a novel combination of documented techniques, leading to improved segmentation accuracy. The Shannon Energy is first used to find sounds of interest. The algorithm subsequently uses the Mel-Scaled WaveleTransform (MSWT) which is a modified Mel-Frequency Cepstral Coefficient (MFCC) algorithm with the Discrete Wavelet Transform (DWT) in order to reduce the impact of noise on the coefficients. The coefficients and sounds of interest are used to distinguish S1 from S2 and segment the signal accordingly. The algorithm is tested on real signals and is compared to a simpler Shannon Energy algorithm and to a traditional MFCC based algorithm. The new algorithm presents an improvement in accuracy especially when signals contain noise. It is therefore less susceptible to outside interference and could be used more accurately in a hospital setting.

INTRODUCTION

Auscultation is a valuable method for the detection of many heart disorders and dysfunctions. There are other diagnostic methods available such as the electrocardiogram (ECG), the echocardiogram and the ultrasound, but heart sound auscultation is the most common one due to its low cost and non-invasive ability to provide information concerning the heart valves and hemodynamics of the heart. The phonocardiogram (PCG) is a recording of the heart sounds and murmurs. It contains the first and second (S1 and S2) sound components associated with the closure of the valves during systole and diastole, as well as any abnormal components. The segmentation of the heart beats in a phonocardiogram is done prior to the analysis of the heart sounds for diagnostic purposes [1-3]. Many different methods of heart sound segmentation and identification have been introduced in the past using techniques such as the wavelet transform [2-4], Shannon energy [1-9], mel-frequency cepstral coefficients (MFCC) [1, 3, 9, 10], and the mel-scaled wavelet transform (MSWT) [10].

The time-domain and frequency-domain features alone were found to be insufficient since the heart sound signals contain non-stationary characteristics. Therefore, the wavelet transform, a time-frequency representation technique, has been proposed to characterize recorded heart sounds, providing information on the time-frequency content of the phonocardiogram during the whole cardiac cycle [10, 11]. The wavelet transform eliminates high frequency noise, but can make false detection for noises overlapping in frequency. Consequently, as an attempt to reduce the false detections, a method combining the wavelet transform and Shannon energy has been proposed to locate the fundamental heart sound lobes which are computed from the low frequency components of the signal [2, 3].

To identify these heart sound components, many authors have proposed a different method based on mel-frequency cepstral coefficient (MFCC). To extract the features from the phonocardiogram signal, MFCC is used, giving good results for clean heart sounds. However, since it is sensitive to the recording frequency response and its performance is not as good in a noisy environment. Therefore, P. Wang *et al.* have proposed to replace the MFCC by the mel-scaled wavelet transform (MSWT) which applies the wavelet transform to the mel spectrum of the phonocardiogram [10]. Their suggested method has produced encouraging results compared with those obtained using the MFCC.

In this paper, the proposed approach is based on using the wavelet transform and Shannon energy techniques mentioned previously in combination with the heart rate estimates from Shannon energy to identify the first heart sound component S1.

In addition, to confirm the S1 results, the combination of MSWT and k-means clustering is executed on the original signal to extract and classify these heart sound components. Finally, the results obtained using these two methods are compared.

METHODOLOGY

Identification of Sounds of Interest (SOI)

It has been suggested in [2] that the PCG signal S(n) can be modeled as

$$S(n) = F(n) + O(n) = F(n) + C(n) + N(n), \qquad (1)$$

where F(n) represents the fundamental components of the heart sounds S1 and S2 (SOI), and O(n) represents other sounds which can be decomposed into C(n), other heart sound components (such as murmurs, etc) and N(n), the noise component.

The first step consists in isolating F(n) by running the signal through an adaptive sublevel tracking module [2]. This module is based on a reiterative process involving wavelet filtering. The 4th order Daubechies wavelet (db4) is used with 7 levels of decomposition. The approximation and detail coefficients are passed through an adaptive threshold. The threshold in the j-level during the kth iteration is defined as

$$Th_{j,k} = |Mean_{j,k}| + f_{j,k} \cdot Std_{j,k}, \qquad (2)$$

where $Mean_{j,k}$ is the mean value of the coefficients, $f_{j,k}$ is an adjustment factor which is varied between 2 and 3, and $Std_{j,k}$ is the standard deviation of the coefficients for a given level and iteration.

The larger signals (coefficients > $Th_{j,k}$) are kept as part of the SOI and the lower signals (coefficients < $Th_{j,k}$) are passed through the wavelet transform again. The minimum likelihood method is used to adjust the stopping criterion S_p.

$$S_p = \left| \frac{E(O_k^2) - E(O_{k-1}^2)}{E(O_k^2)} \right|, \qquad (3)$$

where $E(O_k^2)$ is the expected value of square other signal set *O(n)* at iteration *k* and $E(O_{k-1}^2)$ denotes the expected value at iteration *k-1*.

To extract the envelope of the signal, the Shannon energy is calculated according to

$$E_s = \frac{-1}{N} \sum_{i=1}^{N} H_{norm}^2(i) \cdot \log H_{norm}^2(i), \qquad (4)$$

where,

$$H_{norm} = F(n) / \max(F(n)), \qquad (5)$$

and N is the length of the selected window. This energy calculation emphasizes medium energy components and attenuates low intensity signals compared to high intensity signals [7]. The Shannon energy is calculated by using a 20ms window with a 10ms segment overlap. The significant sounds are then found through the zero crossing points of the normalized Shannon energy

$$E_{s_{norm}} = E_s - \langle E_s \rangle, \qquad (6)$$

where < > represents the mean operator.

The second stage is the identification of SOI based on physiologically inspired properties. To validate the peaks in energy that are found we observe the following properties: sound lobe duration, time interval between sound lobes and the peak energy of the sound lobes. The following criterions are used to identify S1 and S2 [3].

1) All sound lobes have to be between 30ms and 250ms. This is valid for both healthy and cardiac patients. Any sound lobes outside of this range are discarded for further processing.
2) If the time interval between two sounds is smaller than 50ms, it is determined that the sound has been split. The sound with the greatest energy is kept for further processing and the other is discarded.
3) For auscultation done at the apex (5th intercostal space), the energy and length (time duration) of S1 is generally greater than S2. For auscultation in the aortic or pulmonary area the energy of S2 is greater than S1.

A final processing step in the preparation of segmentation is the utilization of the approximate heart rate of the patient to determine expected time intervals. The FFT of the Shannon energy is calculated and filtered using a low-pass filter. The peak frequency of this spectrum provides a useful approximation of the patient’s heart rate. Based on this information and an error margin of +-20%, the interval between S1 sounds is verified and errors due to missing S2 sounds are eliminated or reduced.

The sounds of interest are then compared based on the approximate heart rate intervals, peak energy and expected S1-S2-S1 pattern and identified for segmentation.

Mel-Scaled Wavelet Transform Validation

The mel-scaled cepstral coefficients (MFCC) are readily used in the voice identification field. The mel-scale provides a scaling of the frequency spectrum similar to the human ear's response. Coefficients are then extracted with a mel-scaled filter bank and used to characterize the sound. The dimensionality of the data is reduced by the use of the discrete cosine transform (DCT).

Wang et al. [10] have proposed a mel-scaled wavelet transform (MSWT) that serves the same purpose as the MFCC but helps reduce the impact of noise on the coefficients.

In order to provide a validation of the segmentation results obtained in the first part of the algorithm, the MSWT was implemented and k-means clustering on the coefficients was used to separate S1 from S2 sounds. This information, combined with the results of the first algorithm, results in a more accurate ultimate determination of S1 and S2 and segmentation of the PCG.

The MSWT is implemented by first blocking and windowing the data with a Hamming window. The FFT of each windowed section is then taken and multiplied by the mel-scaled filterbank. The discrete wavelet transform (DWT) is then used to reduce the dimensionality of the data.

The coefficients for each sound lobe identified in the first part of the algorithm are then summed to give a single set of coefficients for each sound of interest. K-means clustering then separates the sounds of interest into two groups by minimizing the Euclidian distance between each set of coefficients and two centroids. This results in a group of S1 sounds and a group of S2 sounds. The results from this step are used to validate the results from the first part of the algorithm.

RESULTS AND DISCUSSION

Testing method

PCG recordings were done at the 5th intercostal spacing on non-cardiac patients as this was deemed the position providing the most precise information for segmentation purposes. These were done at 44.1 kHz sampling rate. Pre-recorded pathological heart sounds were also used to further test the robustness of the algorithm to different pathologies. The results of the segmentation algorithm were then compared to the manually segmented data.

Preliminary results of the MFCC and MSWT algorithms were compared. The k-means clustering technique for differentiating between S1 and S2 sounds was used on the MFCC and MSWT results.

Results

Results of S1 identification for the first part of the algorithm (without MSWT or MFCC) were generally successful as can be seen in Table 1.

Table 1: Statistical Results of S1 Identification from SOI

Samples	**Correct S1 / Correct Other**	**False Positive**	**False Negative**	**Specificity/ Sensitivity**
11 Samples (7 healthy, 4 pathological)	202 / 206	31	33	86.9% / 86.0%

A single pre-recorded example of a mitral regurgitation case significantly lowered the sensitivity and specificity. In this case, S2 sounds were identified as S1 because this pathology makes the S2 sounds longer and more energetic than S1 sounds. This accounts for 27 of the 31 false positives and 29 of the 33 false negatives. Without this sample, sensitivity and specificity are both 98.1%. However, the MSWT correctly identified the S1 and S2 sounds of this particular case. Thus, with the use of the MSWT S1 validation step, this error would have been accurately corrected. In this same case, replacing the MSWT algorithm with the traditional MFCC still provides errors and is thus not useful in correcting the first algorithm.

In figures 1 and 2 below, we can clearly see the attenuating effect of the MSWT on the noise peak located between 1.9 s and 2.1 s when compared to the MFCC. It can also clearly be seen from these figures that for the MSWT, all characteristics provide information about each sound lobe. However, with the MFCC, only one of the characteristics (the blue line in figure 2) provides information about the sound lobes. The other characteristics of the MFCC are similar for all parts of the sound and characterize heavily the noise components of the signal. This provides clear indication of the interest in the MSWT validation step which clearly reduces the impact of noise on the coefficients.

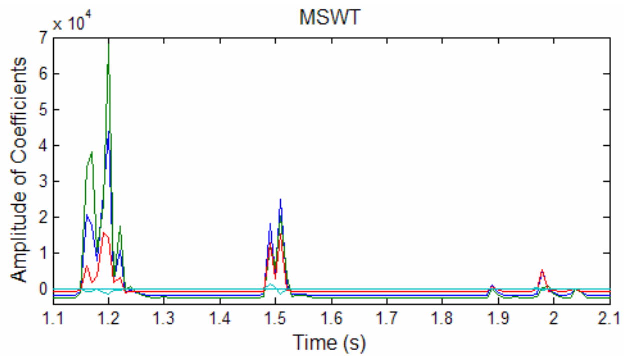

Figure 1: MSWT for single heart beat

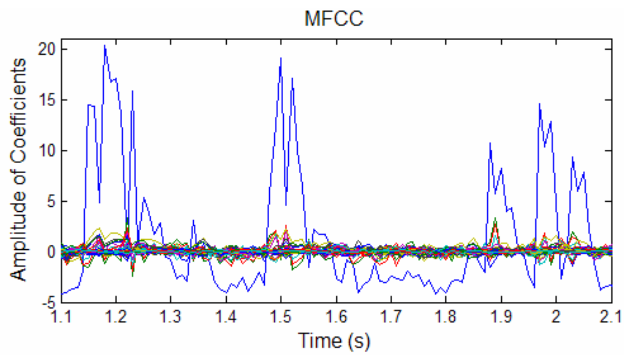

Figure 2: MFCC for single heart beat.

In only one of the samples was the MFCC method slightly more accurate than the MSWT. In 7 of the 11 samples, the MSWT was 100% correct in its distinction of S1 and S2 sounds. This shows that the MSWT is indeed more useful in distinguishing S1 from other sounds than the MFCC in most real life cases such as these where conditions are often not perfect and noise and artifacts do exist.

CONCLUSION

This article proposed the use of an adaptive threshold wavelet transform filtering technique used with Shannon energy, physiological factors and heart rate approximation to properly identify S1 sounds and segment the PCG. However, this method can still present some errors when faced with complex signals. Therefore, the addition of an MSWT validation step was proposed. Preliminary results indicate that the MSWT is less prone to noise than the MFCC and can distinguish S1 sounds from others when faced with complex signals. Future implementation of more robust clustering techniques such as various neural networks have the potential of making the use of the MSWT validation a successful technique to improve accuracy of S1 detection and PCG segmentation.

REFERENCES

[1] S. Chauhan, P. Wang, C. Sing Lim, and V. Anantharaman, "A computer-aided MFCC-based HMM system for automatic auscultation," *Computers in Biology and Medicine,* vol. 38, pp. 221-233, 2008.

[2] P. Wang, Y. Kim, L. H. Ling, and C. B. Soh, "First heart sound detection for phonocardiogram segmentation," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, Shanghai, 2005, pp. 5519-5522.

[3] D. Kumar, P. Carvalho, M. Antunes, P. Gil, J. Henriques, and L. Eugénio, "A new algorithm for detection of S1 and S2 heart sounds," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Toulouse, 2006.

[4] H. Nazeran, "Wavelet-based segmentation and feature extraction of heart sounds for intelligent PDA-based phonocardiography," 2007.

[5] S. Choi and Z. Jiang, "Comparison of envelope extraction algorithms for cardiac sound signal segmentation," *Expert Systems with Applications,* vol. 34, pp. 1056-1069, 2008.

[6] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, L. Eugénio, R. Schmidt, and J. Habetha, "Detection of S1 and S2 heart sounds by high frequency signatures," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, New York, NY, 2006, pp. 1410-1416.

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

[8] A. D. Ricke, R. J. Povinelli, and M. T. Johnson, "Automatic segmentation of heart sound signals using hidden Markov models," in *Computers in Cardiology*, Lyon, 2005, pp. 953-956.

[9] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, "Phonocardiographic signal analysis method using a modified hidden Markov model," *Annals of Biomedical Engineering,* vol. 35, pp. 367-374, 2007.

[10] P. Wang, Y. Kim, and C. B. Soh, "Feature extraction based on mel-scaled wavelet transform for heart sound analysis," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, Shanghai, 2005, pp. 7572-7575.

[11] L. G. Durand and P. Pibarot, "Digital signal processing of the phonocardiogram: Review of the most recent advancements," *Critical Reviews in Biomedical Engineering,* vol. 23, pp. 163-219, 1995.