Detection of the First Heart Sound using a Time-delay Neural Network

T Oskiper, R Watrous

Zargis Medical Corp., Princeton, NJ, USA

Abstract

A method for detecting the first heart sound (S1) using a time-delay neural network (TDNN) is reported. The network consists of a single hidden layer, with time-delay links connecting the hidden units to the time-frequency energy coefficients of a Morlet wavelet decomposition of the input phonocardiogram (PCG) signal. The neural network operates on a 200 msec sliding window with each time-delay hidden unit spanning 100 msec of wavelet data.

Heart sounds were recorded from 30 subjects for 20 seconds at each of four standard auscultatory sites using a commercially available electronic stethoscope. A training set comprised of half of the heartbeats from 20 randomly selected subjects was created. The network was trained on this set and tested on the full data set. The average performance is 1.6% deletion error and 2.2% insertion error. This level of S1 detection is considered satisfactory for analysis of the phonocardiogram signal.

1. Introduction

Computer-assisted auscultation of the heart has the potential to provide a cost-effective technology for screening asymptomatic subjects for valvular and congenital heart disease. Realizing this potential will depend on developing algorithms that can identify and classify basic heart sounds and murmurs accurately and consistently [1,2,3,4]. Such algorithms will need to reliably accommodate the variability across patients of the acoustic manifestation of the fundamental physiological events of the heart cycle, which may arise from normal differences in individual physical parameters or, more critically, differences in health status.

Since the first and second heart sounds serve as acoustic landmarks for ventricular systole/diastole, the invariant recognition of these sounds is particularly important. It has been noted that the time-frequency analysis of the cardiac acoustic signal is particularly appropriate due to the nonstationarity of this signal [5,6].

Detection of the first heart sound is an important step toward the analysis and interpretation of the cardiac acoustic signal, as the first sound identifies the onset of ventricular systole. The ability to detect the first heart sound is particularly important in the absence of a synchronizing ECG signal. Furthermore, the ability to identify the first heart sound without reference to S1-S2 timing information is particularly important for cases of rapid or arrhythmic heartbeats, in which systolic/diastolic intervals may be nearly equal, or highly irregular.

The purpose of this study is to develop an algorithm for the detection of the first heart sound using only the information from the PCG. The algorithm is developed using a database of heart sound recordings obtained from 30 volunteer subjects in good health condition. The heart sounds were recorded for 20 seconds at each of four standard auscultatory sites using each of 2 commercially available electronic stethoscopes. (In this paper, results are reported based on one recording site, specifically the 4th left intercostal space, and one stethoscope.) The stethoscopes were held in place using an elastic bandage, with stabilizing assistance from the subject as needed. The signals were filtered by an anti-aliasing filter with a cutoff frequency at 2 kHz and digitized by a 16-bit A/D converter sampled at 11 kHz. Additional signal modalities, including the electrocardiogram (ECG), lead II, respiration, and noninvasive carotid pressure were also filtered and recorded simultaneously. The time locations of the QRS peaks of the ECG signal are used as fiducial points for selecting S1 signal segments from the PCG signals and also used in determining the neural net objective function and in validation and performance analysis of the algorithm outputs.

2. Methodology

We propose a method based on time-frequency analysis in combination with a time-delay neural network for the detection of the first heart sound in the PCG signal recorded from a human subject.

2.1. Time-frequency analysis

Following the terminology in [7], analytic wavelet transform of a function f at time u and scale s using a wavelet ψ is given by

$$Wf(u,s)=\langle f,\psi_{u,s}\rangle=\int_{-\infty}^{+\infty} f(t)\,\psi_{u,s}^{*}(t)\,dt \tag{1}$$

where

$$\psi_{u,s}(t)=\frac{1}{\sqrt{s}}\,\psi\left(\frac{t-u}{s}\right). \tag{2}$$

From this, the time-frequency energy density i.e., scalogram can be defined as

$$P_W f(u,\xi)=\left|Wf\left(u,\frac{1}{\xi}\right)\right|^2 \tag{3}$$

due to the relation

$$\int_{-\infty}^{+\infty}|f(t)|^2\,dt=\frac{2}{C_\psi}\int_{0}^{+\infty}\int_{-\infty}^{+\infty}P_W f(u,\xi)\,du\,d\xi \tag{4}$$

where C_ψ is a normalization constant, the finiteness of which is called the admissibility condition.

For our study we use the Morlet wavelet (with $\omega_0=5$)

$$\psi(t)=\pi^{-1/4}\left(e^{-i\omega_0 t}-e^{-\omega_0^2/2}\right)e^{-t^2/2} \tag{5}$$

due to its superior time-frequency localization properties.

All scalogram computations are performed in discrete time using Matlab with a 1kHz sample rate in the time axis. The scale axis is sampled using 8 voices per octave for 5 octaves, from which 40 wavelet scales scanning a frequency range of 10Hz-299Hz are obtained. (Scale to frequency conversion is performed based on the fact that Morlet wavelet at a given scale is a Gaussian waveform whose Fourier transform is also Gaussian shaped, from which the equivalent frequency corresponding to a particular scale is chosen as the frequency where the Fourier transform of the wavelet function at that scale reaches its maximum value.) Also to account for the logarithmic sampling in scale axis, the continuous scalogram formula (3) is normalized properly for each voice.

2.2. TDNN architecture

The time-delay neural network is applied since it provides with a shift-invariant detector that enables the time-localization of S1 with reference to the network output [9]. Magnitude square (i.e. scalogram) of the continuous wavelet transform based on complex Morlet wavelet function is used as the input. The input layer for the two-layer network consists of 40x200 array of scalogram coefficients and the second layer contains five hidden units. Each hidden unit has a time span of 100 msec and the hidden layer slides over the 200 msec input window in 10msec increments. The output unit combines all the activations from the hidden layer over this 200msec window. So the network is presented with 200msec of PCG scalogram data sequentially in time. (The network window also moves over the scalogram in 10msec increments in time which is achieved by shifting the input layer 10msec forward in time over the scalogram data.)

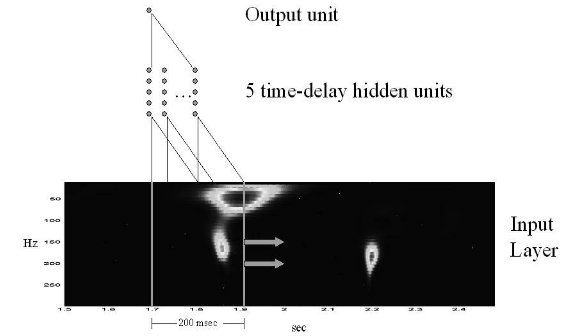

Figure 1. Illustration of the TDNN architecture

2.3. Training of the TDNN

The training set is comprised of half of the heartbeats from 20 randomly selected PCG signals. In this study, a two pass training approach is used on this set. In the first pass, the neural net is trained to respond to S1 (output=1) and not to respond to non-S1 regions (output=0) except that its behavior for the second heart sound (S2) is not penalized. And in the second pass the first network is further trained and is forced to suppress its output for S2 while still responding to S1.

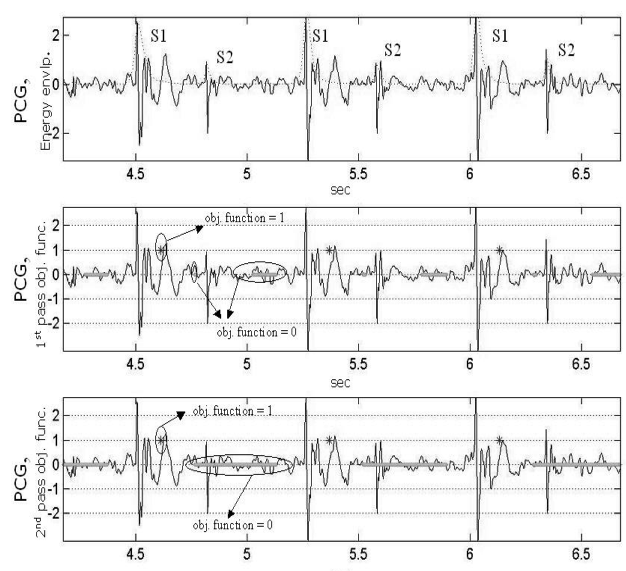

Figure 2. Illustration of the objective function values for 1st and 2nd pass training.

The peak of the QRS complex of the ECG signal is used as the fiducial point to parse the PCG into separate cardiac cycles chosen as the R-R intervals. The objective function for the first phase of training is defined as follows. For each cardiac cycle the energy envelogram of the PCG signal is computed and the time instant at which it achieves its maximum value within 200 msec of the QRS peak is noted and this portion of the PCG is called the S1 segment. The objective function at this time instant is assigned the value 1. (This is the time instant at which maximum energy of the S1 sound lies in the center of the network's time span.) A buffer of sufficient duration is used right before and after this time instant in which the network is not penalized (the gradients and errors are weighted by zero) for its output. The non-S1 region is determined as the remaining of the cardiac cycle excluding the S1 segment and the above mentioned buffers. During the first pass training phase, the network is required not to respond (objective function is equal to zero) in the non-S1 region except for those parts of the non-S1 region that includes substantial energy of the second heart sound (S2) for which it is not penalized. (All the errors and gradients are weighted by zero for such inputs.) So for each cardiac cycle the network is presented with a single S1 segment and a stream of non-S (no S1 and S2) segments obtained from the non-S region in 10msec increments. Finally, all network parameter updates are performed after each cardiac cycle using stochastic gradient descent method in which the gradient and error for S1 segment are proportionally weighted by the number of non-S1 segments the network has seen in that heart cycle.

After the network converges, 1^{st}-pass network coefficients are used as the initial condition for the second phase of the training. In the second pass, the network is required to go over the whole cardiac cycle and respond to S1 while not responding during the whole non-S1 region, i.e., the output over the S2 region is required to be zero as well. At the end of the second phase training the converged network is used for the detection of S1.

3. Examples

In the following figures, we show the response of the TDNN after 1^{st} and 2^{nd} training phases for some example PCG signals in the database. The corresponding ECG signal is also plotted in dotted lines for each case and is used for validation and performance evaluation. For each figure, the top plot shows 10 seconds of PCG signal. The TDNN obtained after 1^{st}-pass training is applied to this signal and its output is plotted in the middle plot. At this stage the network is observed to respond to both S1 and S2 as expected. On the other hand, the final TDNN whose output is plotted in the bottom plot of each figure is observed to respond to S1 while at the same time successfully suppressing its responses to S2.

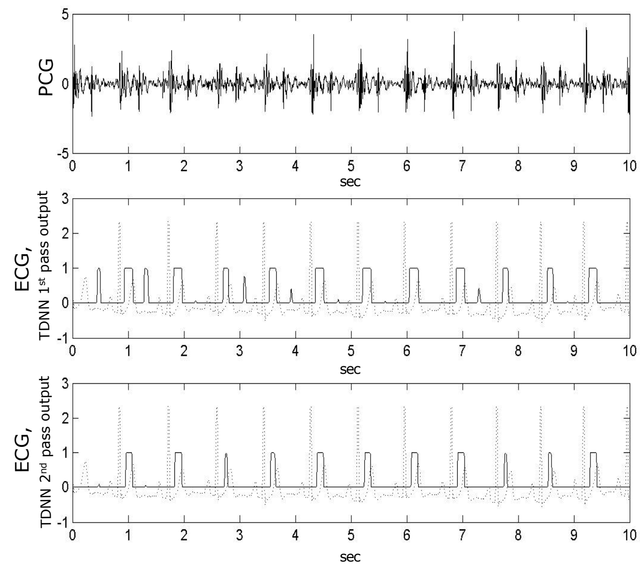

Figure 3. Sample PCG (10 sec) and the TDNN outputs (solid) after 1^{st} and 2^{nd} pass training overlaid with ECG signal (dotted).

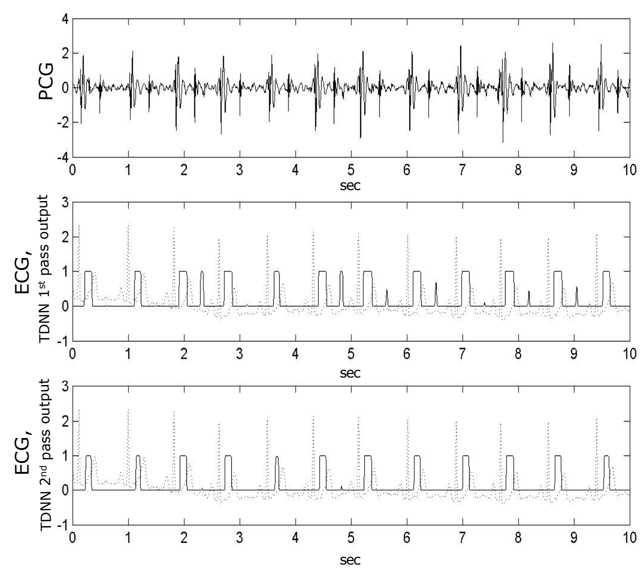

Figure 4. Sample PCG (10 sec) and the TDNN outputs (solid) after 1^{st} and 2^{nd} pass training overlaid with ECG signal (dotted).

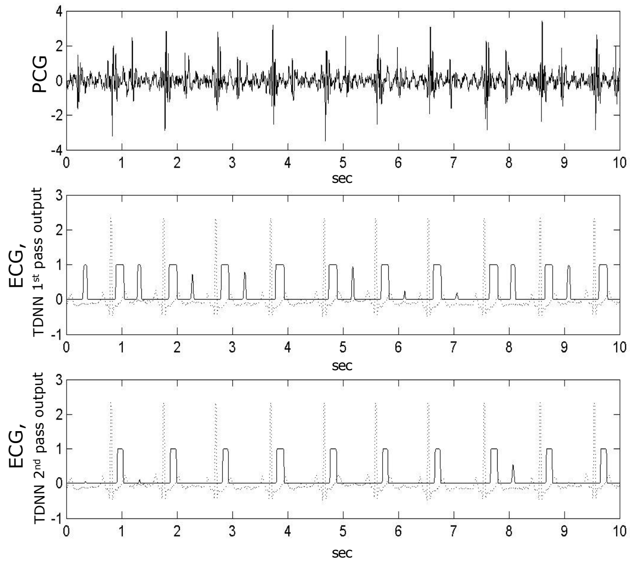

Figure 5. Sample PCG (10 sec) and the TDNN outputs (solid) after 1st and 2nd pass training overlaid with ECG signal (dotted).

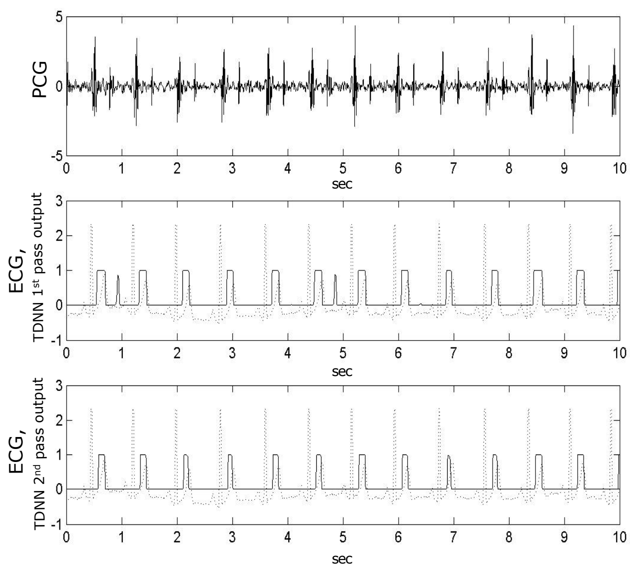

Figure 6. Sample PCG (10 sec) and the TDNN outputs (solid) after 1st and 2nd pass training overlaid with ECG signal (dotted).

4. Results and conclusion

After the completion of both phases of the training, the resulting neural network is tested on all the data. The time instants at which the rising edge of the TDNN firings cross a selected threshold level of 0.5 are determined and noted as the detection points for the first heart sound declared by the algorithm. At the end of the first pass training, 0.3% deletion error and 27% insertion error performance is obtained on all 30 PCG signals. The vast majority of the insertion errors at this stage are observed to be due to the second heart sound. After the second phase of the training, overall performance for all 30 subjects resulted in 1.6% deletion errors and 2.2% insertion errors. As expected, the number of insertion errors decreased significantly while the amount of deletion errors remained at a sufficiently low level. In conclusion, this level of performance for S1 detection is considered to be satisfactory for analysis of the phonocardiogram signal.

References

[1] Liang H, Lukkarinen S, Hartimo I. Heart sound segmentation algorithm based on heart sound envelogram. Computers in Cardiology 1997;24:105-108.

[2] Liang H, Lukkarinen S, Hartimo, I. A heart sound segmentation algorithm using wavelet decomposition and reconstruction. EMBS 1997:1630-1633.

[3] Reed TR, Reed NE, Fritzson P. The analysis of heart sounds for symptom detection and machine-aided diagnosis. Proceedings of the 4th International EUROSIM Congress 2001.

[4] Rajan S, Doraiswami R, Stevenson M, Watrous RL. Wavelet based bank of correlator approach for phonocardiogram signal classification. Proc IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis 1998:77-80.

[5] Chen D, Durand LG, Lee HC. Time-frequency analysis of the first heart sound. Part 1: Simulation and analysis. Med Biol Eng Comput 1997;35:306-310.

[6] Wood JC, Buda AJ, Barry DT. Time-frequency transforms: a new approach to first heart sound frequency dynamics. IEEE Trans Biomed Eng 1992;39:730-40.

[7] Mallat S. A Wavelet Tour of Signal Processing, 2nd ed. London: Academic Press, 1999.

[8] Yoganathan AP, Gupta R, Udwadia FE, Miller JW, Corcoran WH, Sarma D, Johnson JL, Bing RJ. Use of the fast Fourier transform for frequency analysis of the first heart sound in normal man. Med Biol Eng 1976;14:69-73.

[9] Waibel A, Hanazawa T, Hinton G, Shikano K, Lang K. Phoneme recognition using time-delay neural networks. IEEE Acoustics Speech and Signal Processing 1989;37: 328-339.

Address for correspondence.

Zargis Medical Corp.
755 College Road East
Princeton, NJ 08540
E-mail: {toskiper,watrous}@zargis.com