# Whole Fundamental Heart Sound ANN-based Detection using Simple Features

Ana M. Gavrovska, *Member, IEEE*, Milorad P. Paskaš, *Member, IEEE*, Dragi M. Dujković and Irini S. Reljin, *Senior Member, IEEE*

***Abstract*** **— The paper considers detection of whole fundamental heart sounds (S1 and S2 heart sound) using simple feature vector and having annotated database. Determination of the precise beginning and end of the heart sound in time domain is of great importance, especially in the cases of split existence and/or murmur and heart events existence next to fundamental heart sounds. About 99% of accuracy is obtained in the simulation presented in this work.**

***Keywords*** **— ANN, STFT, PCG, wavelet decomposition, feature vector, regression.**

## I. Introduction

ARTIFICIAL neural networks (ANNs) and generally artificial intelligence (AI) are capable to resolve numerous difficult tasks introducing some level of intelligence. ANN has the ability to collect knowledge from experimental data and learn the functional relations between the input and appropriate output changes. In the paper this ability is tested on phonocardiograms (PCGs).

Heart sounds are auscultated and monitored through visual representations assisting physicians in making decisions concerning diagnosis. Their auscultative accuracy rate is often low [1] and misleading heart event recognition may be interpret wrong in the following making decisions concerning patient's treatment continuing [2-4].

Phonocardio signal has two fundamental heart sounds, S1 and S2, that determines systole and diastole periods of the cardiac cycle. Their determination is crucial for defining the heartbeat in recorded PCG without any new realizations of familiar reference signals, as well as for other heart sounds, events and/or murmurs recognition.

A. M. Gavrovska, PhD stipendiary of the Ministry of Science and Technological Development, Republic of Serbia, is with the Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia; (phone: 381-11-3370143; e-mail: anaga777@gmail.com).

M. P. Paskaš is with the Innovation Center of the Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia; (phone: 381-11-3370143; e-mail: milorad.paskas@gmail.com).

D. M. Dujković is with the Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia; (e-mail: dragi@etf.rs).

I. S. Reljin is with the Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia; (e-mail: irinitms@gmail.com).

Sufficiently precise assigning time positions to limitations of S1 and S2 heart sound is of importance. Cases where S1 and/or S2 splits, adjacent murmurs or adjacent heart events (e.g. ejection click) exist represent the cases where such assigning is required.

This paper investigates the possibility of having fully, manually annotated database (in this case for S1 and S2) for segmentation and detection purpose. It considers the accuracy of detecting whole fundamental heart sounds, meaning implementing simple ANN and simple feature vector in order to precisely segment them in time.

The paper is organized as follows. In Section II the simulation is presented. Different types of trained and tested signals are used. The results made by using simple feature vector of different length and implementing ANN can be found in Section III. Several concluding remarks and observations on made consideration are derived in Section IV.

## II. Simulation

Signals of different quality that were used in this paper can be found in [5] and appropriate used signal type percent diagram is presented in Fig.1. The chosen test-bed does not possess high-energy and/or low-frequency murmurs, and these cases should be investigated in future work.

The system should be able to precisely recognize time limitations of S1 and S2 heart sound. Making decision about time positions of the whole S1s and S2s can be misleaded by having different heart events in close time proximity, like ejection clicks (ECs), opening snaps (OSs), even heart sounds (S3 and S4), etc. Split existence in S1 and S2 is one of the most common examples that can lead to their misrepresentations.

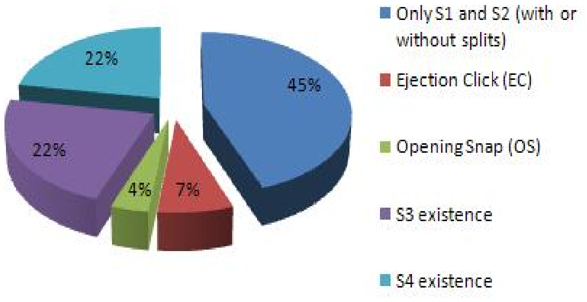


Fig. 1. Signal type percent diagram.

Time position and duration of fundamental heart sounds, S1 and S2, were analyzed in this work and they were manually segmented in the test-bed. These manual annotations provide appropriate target values.

Using markers and feature vectors it is possible to train the system in order to make the segmentation automatic and/or semiautomatic. Even previously made analythical segmentation of the S1 and S2 heart sounds may be improved using the neural network tool for fitting purpose. This can be an important part of one PCG signal processing module (SPM) [6-7].

### A. Feature vector and marker selection

For simulation purpose, only one two-second segment was chosen randomly from each signal-patient, which statistics was presented in Fig.1. In such way the number of heartbeats is one to six in each segment. Segments are decimated to frequency of 2kHz.

The common way of low-cost cardiac functionality monitoring is using time-frequency representations of PCG, like spectrograms. In Fig.2. two-second segment from healthy patient is depicted in time and time-frequency domain.

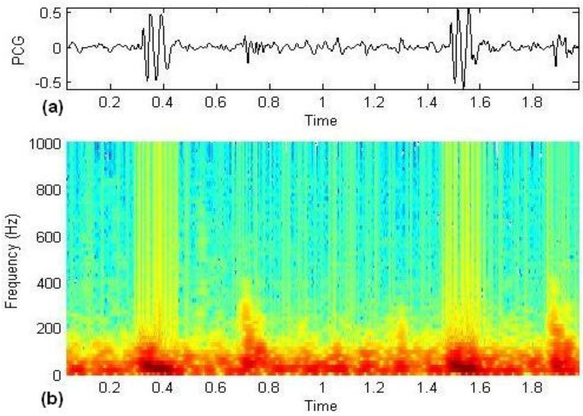


Fig. 2. Two-second segment from healthy patient signal in (a)time and (b)time-frequency domain (spectrogram).

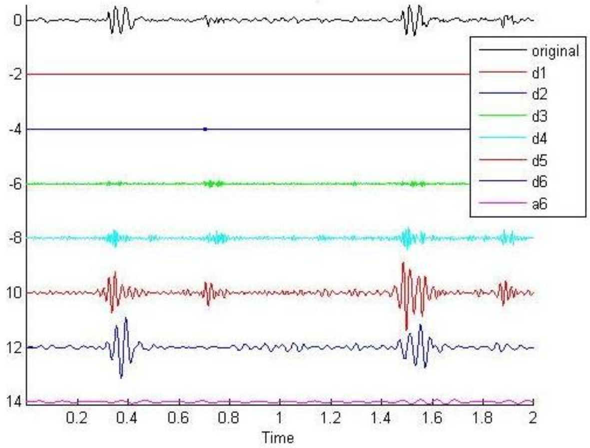


Fig. 3. DWT decomposition of the selected segment (level 6).

The first part of feature vector (FV) is obtained using discrete wavelet decomposition (DWT) of the fifth level (db21). Wavelets are common used tool in PCG analysis. Detail (*d1 - d6*) and approximation coefficients (*a6*) for DWT decomposition of level six [6] of the selected segment are presented in Fig.3. Only signal *d5* is used in this work.

The STAE (Short-Time Average Energy) [8] of decomposed signal detail coefficients *d5* is calculated as:

$$E_n = \sum_{m=-\infty}^{+\infty} \left(d_5(m)W(n-m)\right)^2, \qquad (1)$$

where $W(n)$ represents Hamming window function of length $n$=128. In Fig.4. STAE of *d5*, $E_n$, is presented with the original (decimated) signal-segment.

Linear spaced $N$ points were chosen from normalized STAE signal for constructing the first part of the feature vector. Local maximums of the obtained signal are depicted as markers in the Fig.4.

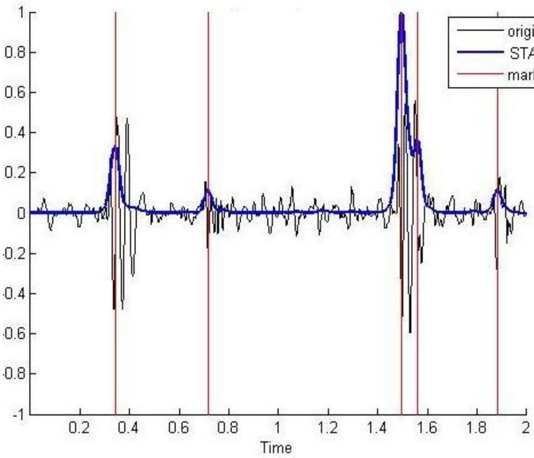


Fig. 4. STAE of *d5* coefficients with markers as local maximums.

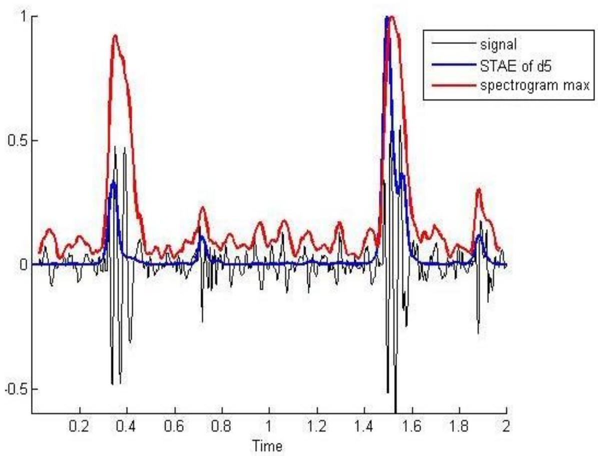


Fig. 5. STAE of *d5* coefficients and maximums of spectrogram columns.

The other part of the feature vector is obtained using STFT (Short-Time Fourier Transform) spectrograms

(Fig.2.(b)) and calculating maximums of each spectrogram column. The final signal is normalized as well. Spectrograms are generated using windows of 128 samples with 120-sample overlapping and the markers represent local maximums of the final signal. Linearly spaced *N* points were chosen.

Both the STAE of d5 coefficients (blue line) and maximums of spectrogram columns (red line) are depicted in Fig.5. These signals are chosen to be fundamental for FV constructing.

The selected feature vector has 2*N+1 elements, consisting of *N* elements using DWT, *N* elements using STFT and one element representing the marker. The marker can be manually or automatically selected and represent the point where the heart sound is surely situated. For simulation purposes we used local maximums of both the first and the second part of the feature vector before selecting linearly spaced *N* elements in both cases. The feature vector/input and target/output content is presented in Fig. 6.

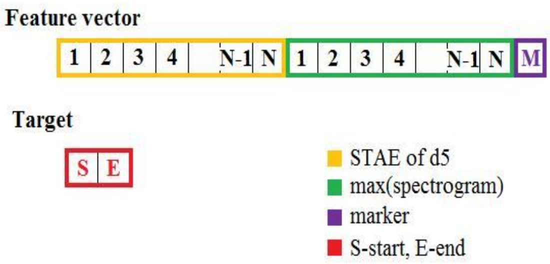


Fig. 6. FV and target content.

Two-element target represent the beginning and the end of each S1 and S2 within the segment, where the marker (last element in feature vector) is located between. Imported values are given in percentage within the two-second segment, so S, E and M have values between 0-100.

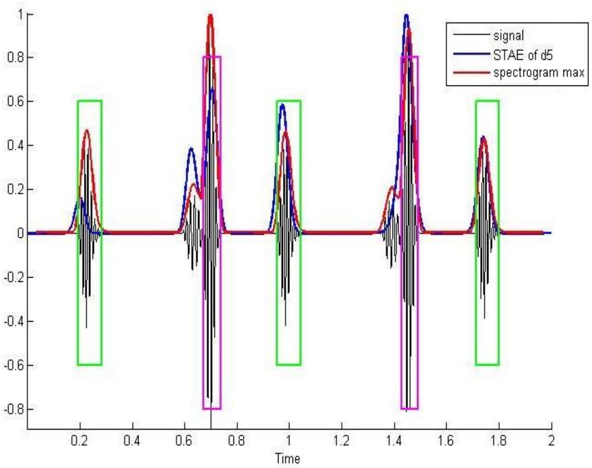


Fig. 7. STAE of d5 coefficients and maximums of spectrogram columns with appropriate targets (magenta (S1) and green (S2) rectangles) for S4 existence case.

In Fig.7. STAE of d5 coefficients and maximums of spectrogram columns are presented with appropriate targets for S4 existence case. In the second S1 within the segment one can see the importance of having maximums of spectrogram columns beside the wavelet coefficients.

### *B. Feed-forward network and performance measures*

The standard two-layer feed-forward network trained with Levenberg-Marquardt was used for fitting purpose. The block-scheme of such network is presented in Fig.8. The fixed number of neurons in the output layer is two.

Different values of *N* as well as different number of neurons in hidden layer were tested. Obtained average iteration number, MSE (Mean Squared Error) and R-values are calculated. MSE measures the performance and represent the average squared difference between outputs and targets. R-values represent the correlation between outputs and targets.

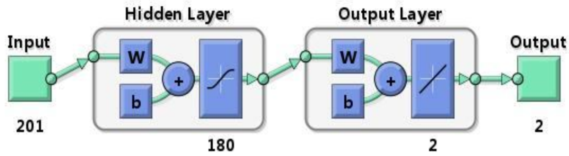


Fig. 8. Two-layer feed-forward neural network for fitting purpose for 201 samples long FV and 180 neurons in hidden layer.

From the number of available vectors 75%, 15% and 15% is used for training, validation and testing set, respectively.

## III. Simulation Results

Linear regression is performed between the network outputs and the corresponding targets. Results for different values of length of feature vector (2*N+1) can be found in Table 1. Correlation (R-) values for training, validation and testing, as well as the overall correlation, MSE and number of iterations are presented for several cases.

Correlation coefficient (R-value) between the outputs and manually obtained targets is a measure of how well the variation in the output is explained by the given targets. The correlation is better as it is closer to 1. The R-value (Table 1) is at least 0.98 for the total response. Results of fitting are accurate, and it is possible to provide accuracy of over 0.99 for the overall R-value.

The further improving accuracy of segmenting fundamental heart sounds and investigation and detection of other heart events can be based on the results made here. The system was able to have around 99% accuracy in detecting these limitations in time domain.

The right selection of the ANN characteristics depends on the quality measure that we choose. In other words, it depends on e.g. what is more important: to have less MSE or to have less iterations and processing time in such algorithm step.

TABLE 1: NEURAL NETWORK FITTING RESULTS.

| Length of FV | Number of available vectors | Number of neurons in hidden layer | R-value (training, 75%) | R-value (validation, 15%) | R-value (testing, 15%) | MSE | Number of iterations | R-value (final) |
|---|---|---|---|---|---|---|---|---|
| 61 | 281 | 100 | 0.99724 | 0.98122 | 0.99012 | 4.31 | 75 | 0.99360 |
| | | 180 | 0.99902 | 0.99791 | 0.99680 | 1.63 | 111 | 0.99854 |
| 121 | 256 | 100 | 0.98420 | 0.98128 | 0.97925 | 25.7 | 38 | 0.98281 |
| | | 180 | 0.99903 | 0.99703 | 0.99453 | 1.53 | 110 | 0.99817 |
| 201 | 258 | 100 | 0.99579 | 0.99467 | 0.99274 | 7.03 | 64 | 0.99531 |
| | | 180 | 0.98564 | 0.97427 | 0.98548 | 24.8 | 39 | 0.98386 |
| 401 | 260 | 180 | 0.99867 | 0.99670 | 0.99551 | 2.25 | 103 | 0.99807 |
| | | 200 | 0.99724 | 0.98122 | 0.99012 | 4.42 | 47 | 0.99360 |

# I. IV. CONCLUSION

Combining ANN with analytical methods, that are common in finding the S1 and S2, it is expected to have satisfactory precise determination of their time intervals. In this case automatically detected markers were used, and having appropriate heart event annotated database, ANN results were satisfactory. The further work need to be dedicated to test-bed extension as well as considering different diagnosis and murmurs (late, early, holo-) adjacent to fundamental heart sounds.

In this work, around 99% of accuracy is provided in detecting the beginning and end of the fundamental heart sounds. The similar principle can be used for some other heart events and heart sounds (S3, S4).

One SPM can be consisted of several parts that are ANN-based for classifying, detection, processing, etc. After making initial classification results, SPM module can use a subsystem as explained for improving event segmentation in time domain. These results are of the great importance in the subsequent sections of SPM, from improving time-frequency determination/segmentation results to decreasing previously made detection and classification errors.

# REFERENCES


[1] S.L. Strunic, F. Rios-Gutierrez, R. Alba-Flores, G. Nordehn, S. Burns, "Detection and Classification of Cardiac Murmurs using Segmentation Techniques and Artificial Neural Networks", in *Proc. of the 20th International Florida Artificial Intelligence Research Society Conference*, FLAIRS-20, pp. 128-133, May 7-9, 2007, Key West, Florida, USA.

[2] C.G. DeGroff, S. Bhatikar, J. Hertzberg, R. Shandas, L. Valdes-Cruz, R. Mahajan, "Artificial Neural Network-Based Method of Screening Heart Murmurs in Children", *Circulation 2001*;103:2711-2716, American Heart Association, Inc.

[3] S.R. Bhatikar, C. DeGroff, R.L. Mahajan, "A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics", *Artificial Intelligence in Medicine* (2005) 33, pp. 251-260.

[4] A. Noponen, S. Lukkarinen, A. Angerla, R. Sepponen, "Phono-spectrographic analysis of heart murmur in children", *BMC Pediatrics* 2007, 7:23, BioMed Central Ltd.

[5] *E-chocardiography Journal* - Cardiac Auscultation. Available: http://rwjms1.umdnj.edu/shindler/auscultation.html.

[6] F.Javed, P.A.Venkatachalam, A.Fadzil, "A Signal Processing Module for the Analysis of Heart Sounds and Heart Murmurs", *Journal of Physics*, *International MEMS Conference*, Series 34, 2006, pp. 1098-1105.

[7] A.M. Gavrovska, M.P. Paskaš, D. Dujković, and I.S. Reljin, "Region-based Phonocardiogram Event Segmentation in Spectrogram Image", 10th Symposium on Neural Network Applications in Electrical Engineering, NEUREL-2010, Faculty of Electrical Engineering, University of Belgrade, Serbia, September 23-25, 2010.

[8] L.R. Rabiner, R.W. Scafner, Digital Processing of Speech Signals, Prentice Hall Signal Processing Series, 1978, pp. 116-135.