World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:17, No:10, 2023

Robust Heart Sounds Segmentation Based on the Variation of the Phonocardiogram Curve Length

Mecheri Zeid Belmecheri, Maamar Ahfir, Izzet Kale

***Abstract*—**Automatic cardiac auscultation is still a subject of research in order to establish an objective diagnosis. Recorded heart sounds as Phonocardiogram (PCG) signals can be used for automatic segmentation into components that have clinical meanings. These are the first sound, S1, the second sound, S2, and the systolic and diastolic components, respectively. In this paper, an automatic method is proposed for the robust segmentation of heart sounds. This method is based on calculating an intermediate sawtooth-shaped signal from the length variation of the recorded PCG signal in the time domain and, using its positive derivative function that is a binary signal in training a Recurrent Neural Network (RNN). Results obtained in the context of a large database of recorded PCGs with their simultaneously recorded Electrocardiograms (ECGs) from different patients in clinical settings, including normal and abnormal subjects, show on average a segmentation testing performance average of 76% sensitivity and 94% specificity.

***Keywords*—**Heart sounds, PCG segmentation, event detection, Recurrent Neural Networks, PCG curve length.

I. Introduction

HEART sounds give indications to the state of the heart and its functioning (rate, rhythm, Fundamental Heart Sounds (FHS), gallops and murmurs). Cardiac auscultation is the first step of physical examination which makes possible the detection of the first signs of heart diseases. Because heart diseases are among the leading causes of death in the world [1], early detection should be considered which can help in stopping their progression.

Since the invention of the stethoscope in 1816 by Dr. Laennec and the description of auscultatory semiology in 1819 [2] in the Treaty of Mediate auscultation, and up till the last decade, cardiac auscultation has not developed much. Research on the physical characterization of physiological and pathological cardiac sounds has not drawn a lot of attention to the research community. However, in clinical practice, the ability to distinguish normal from abnormal sounds (murmur, gallops) remains critical for diagnosis and medical interpretation.

The classical auscultation based on subjective criteria is far from being rigorous for systematic cardiac sounds classification. Indeed, the practitioners use only their auditory perception and past memories of pathological sounds which may lead to errors in practice.

Mecheri Zeid Belmecheri is with Department of Electronics, University of Laghouat, Algeria.

Maamar Ahfir is with Department of Informatics, University of Médéa, Algeria (e-mail: ahfirmaamar@gmail.com).

Izzet Kale is with School of Computer Science and Engineering, University of Westminster, London, UK.

Over the last two decades, auscultation has seen remarkable progress in the areas of enhancement of the signal acquisition and also innovation in the signal processing and analysis of auscultatory signals. These improvements should provide better sensitivity and specificity in the auscultation results.

Heart sounds recorded and represented as PCG signals can be used for automatic segmentation. PCG signal segmentation aims to determine the boundaries of the FHS from the temporal segments representing the cardiac cycles of a heart functioning. It is considered as the most difficult step in heart sound analysis due to interferences from murmurs and other noises, which result in extra-peaks. PCG signal segmentation has been the subject of numerous studies, particularly PCG energy envelope-based methods which have less complexity for implementation, such as in [3]-[6]. In [3], different envelopes such as: Shannon energy, Hilbert transform and cardiac sound characteristic waveform were extracted from a PCG and compared. From all the envelopes the FHSs are located using both fixed and adaptive thresholding used in a clustering technique. In [4], a method based on the second derivative function of the multiscale Hilbert envelope of the PCG was proposed, where each FHS is located between two consecutive maxima of the second derivative function. An adaptive thresholding on the continuous energy signal of the PCG [5] based on the value of the heart rate was proposed in [6]. This method has given promising results on PCGs even with extra-peaks, but with stable normal heart rhythms (i.e., the cardiac cycles/periods are almost identical). However, in the case of PCGs with unstable cardiac rhythms (known as arrhythmias), localization of the FHSs may be inaccurate, which may result in a false interpretation of PCGs.

In this paper, an automatic segmentation method is proposed in order to overcome the problems cited above, and find the accurate localization of the FHSs. This method is based on calculating an intermediate sawtooth-shaped signal from the length variation of the PCG signal in the time domain, and using its positive derivative function in training a Layer Recurrent Neural Network.

The paper is organized as follows: Section II describes the proposed method. Section III describes Layer Recurrent Neural Network used for training. Results and discussion are given in Section IV and Section V concludes the paper.

II. Proposed Method

A. Normalization

Given different recording conditions, the appearance of extra-peaks in a PCG will cause attenuation of the normalized signal if the normalization is done on the maximum value of

the PCG. If the elimination of the extra-peaks that exceed three times the average value of the peaks in an interval of 500 ms as in [6] is deployed, this elimination has led to the elimination of FHSs if the extra-peaks coincide with true FHSs.

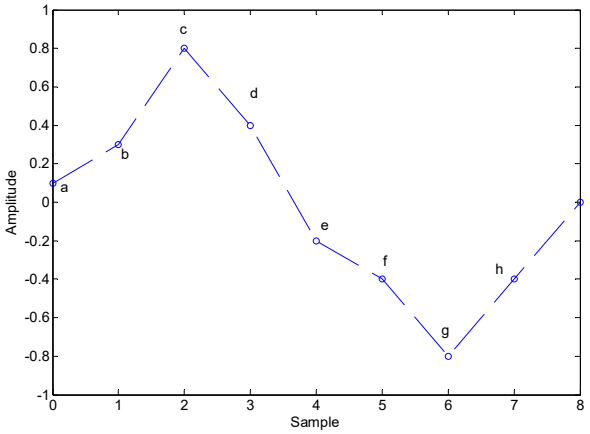

Fig. 1 Calculation concept of a curve length variation: a,b,c ….,h are successive points of a PCG curve from which at each instant the sum of cumulative Euclidean distances between them is calculated

In this paper we propose to normalize the recorded PCG to its absolute maximum value in a time window of 1 second, because in this time window at least one FHS is present [7]. In case where an extra-peak appears in this time window, the PCG in this time window is normalized to this extra-peak. So only the window containing the extra-peak is normalized to the wrong value and is then eliminated and not considered. See an example in Fig. 2.

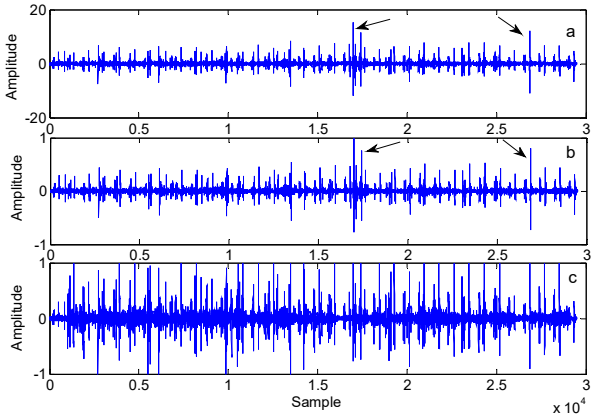

Fig. 2 Elimination of the extra-peaks from the PCG: (a) PCG recorded with two artifacts shown by arrows, (b) PCG attenuated by the normalization to the absolute maximum value and (c) PCG normalized on a 1 second time window

B. Curve Length Variation

After the normalization process, the PCGs are filtered by a band pass filter, where only the frequencies between 30 Hz and 150 Hz are kept. In this band, only the FHS components are present, and other sound components, such as murmurs and noise are filtered out [8].

The Euclidean distance between two points a and b with their coordinates (x_1,y_1) and (x_2,y_2) respectively is defined by the formula:

$$dist(a,b)=\sqrt{(y_2-y_2)^2+(x_2-x_2)^2}. \quad (1)$$

To show how the length variation at each instant is calculated, let us consider the a,b,c,d,…, h successive points of a curve with their respective abscissas 0,1,2,3,4…,7 as shown in Fig. 1.

- For sample *b*, *D(1)=dist (a,b).*
- For sample *c*, *D(2)=dist (a,b) + dist(b,c).*
- For sample *d*, *D(3)=dist (a,b) +… dist (c,d)*
- For the last sample *h*, *D(7)=dist(a,b)+ dist (b,c) + dist (c,d) +….dist(g,h).*

This can be represented for a PCG curve by:

$$\boldsymbol{D(n)=\sum_{k=1}^{n} dist(PCG(k),PCG(k+1))} \quad (2)$$

where *dist (PCG (k), PCG (k+ 1))* is the Euclidean distance between two successive points of the PCG curve where *n* is its sample instant.

Since the function *D* is a sum of the Euclidean distances which are by definition positive, it is always increasing. This can be expressed by a regression function as follows:

$$T(n)=D(n)-Y, \quad (3)$$

where *Y* is the regression function given by:

$$Y=a*n+b\ , \quad (4)$$

with *a* the slope of the regression function given by:

$$a=\frac{cov(D,n)}{var(n)} \quad (5)$$

and *cov(D,n)* is the covariance between *D(n)* and *n*, and *var(n)* is the variance of *n*.

The continuous value *b* is given by:

$$b=\overline{D}-a*\overline{n}, \quad (6)$$

where $\overline{D}$ and $\overline{n}$ are the mean values for *D* and *n*.

Fig. 3 shows results for an example of a normal PCG. The Curve *T*(n) takes the form of a sawtooth function, where the intervals of the rising edges represent the fast changes in the length of the PCG curve in the S1 or S2 (FHS) intervals. The intervals of the falling edges represent the slow variations of the PCG curve length in the intervals of systolic and diastolic silence. The rising edge of the *T* function is localized with the positive part of the signum function of the derivative function *dT(n)/dn*. Every positive part of this binary signal is taken as an FHS location (Fig. 3 (d)).

After localizing the rising edges, and as it is known the systolic duration is less than the diastolic one (i.e. regular PCG) [7], the FHSs are labeled as S1 or S2 according to the

durations between them.

The problem that is faced in this method is the difficulty to locate the right FHSs on a PCG which has close systolic and diastolic durations, diastolic duration less than the systolic one (i.e. irregular PCG) and additive murmurs or noise.

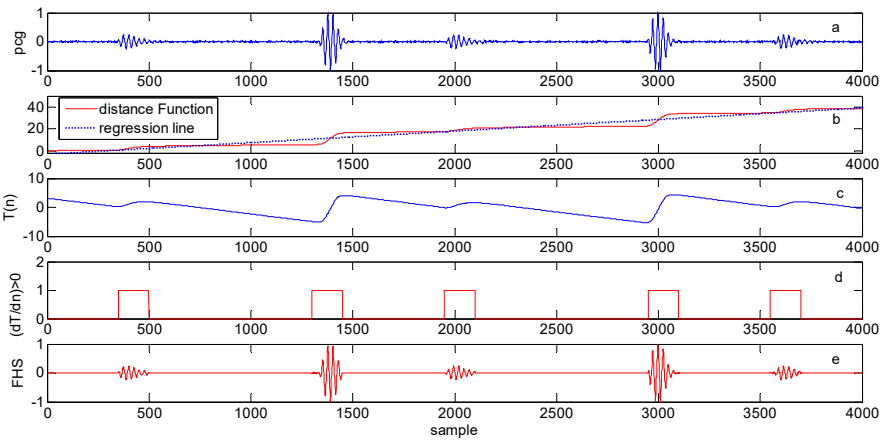

Fig. 3 Results for an example of a normal PCG

III. Layer Recurrent Neural Networks

To overcome the problems of false FHS localization on the positive derivative *dT(n)/dn* function (Fig. 3) and inaccurate identification of the S1 and S2 sounds, a RNN architecture is used. These neural networks are similar to feed-forward networks, except that connections are added between the output of the hidden layers and the input layer in order to use the predicted data at the previous positions in a sequence as input to the network. The outputs of the hidden layers are added to the input layer as contextual information.

The dynamic neural network used in this study is the Layer Recurrent Neural Network (LRNN) [9]. In [10] Elman proposed the first version of this network. The "layrecnet" command in MATLAB generalizes the Elman network to have an arbitrary number of layers and transfer functions in each layer [11].

In order to allow the network to have an infinite dynamic response to the time series input data, a flow loop, with a single delay around each layer of the network is added, except the last layer, as shown in Fig. 4 [11]. The selection of appropriate neural network parameters such as the number of neurons in each layer, the number of hidden layers, and the transfer function types were the most important network parameters considered for this architecture [12].

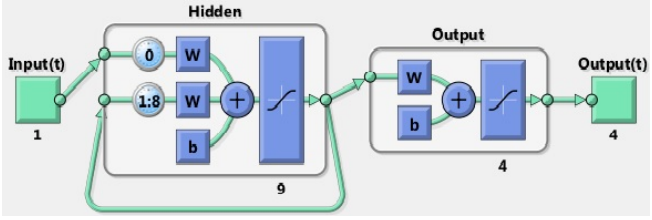

Fig. 4 Designed LRNN composed with: 1 input unit,4 neurons in 4 output layers and 9 neurons in the hidden layer

A. Labeling PCGs by S1 and S2

In order to train the LRNN, a database of PCGs labeled by their S1 and S2 built by [13] has been used. This database is composed of 792 heart sounds recorded from various locations on the chest of 135 patients with multiple recordings per patient. Each PCG has its own annotations of R peaks (which corresponds to the depolarization and contraction of the ventricles, right and left) and end T waves (corresponding to the depolarization and contraction of the atria, right and left) [7]. These annotations were derived from the ECGs recorded simultaneously with the corresponding PCGs.

For labeling a PCG with S1 and S2, in [6] a method is used to localize them from R peak and end T wave annotations. The beginning of the S1 sound is localized by the R peak of the ECG [14], and the duration of S1 is the mean of S1's durations as proposed in [6]. The beginning of the S2 is localized at the end of the T wave on the ECG [14], and the S2 sound is deduced by localizing the center defined by the maximum peak in the Hilbert envelope of the PCG. The S2 duration is defined by the mean of S2 durations as proposed in [6].

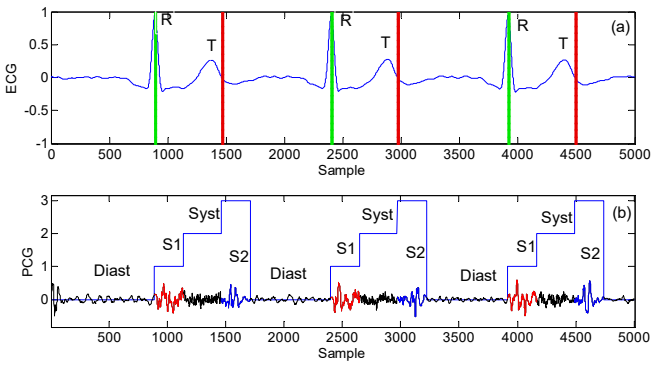

Fig. 5 Recording of a normal heart sound with corresponding ECG tracing; S1 is the first heart sound which marks the beginning of the systole, S2 is the second heart sound which marks the beginning of the diastole

Finally, the S1 sound can be labeled by State 1, the systolic silence by State 2, the S2 sound by State 3 and the diastolic silence by State 0, as shown in Fig. 5.

IV. Results and Discussion

The layer recurrent network proposed for this study is composed by: 1 input unit, 9 neurons in the hidden layer and 4 neurons in the output layer, one neuron to represent each state. The network has a feed-back loop with a tap of 8 delays. Tan-sigmoid transfer functions are used for the hidden and output layers (Fig. 4). 1000 iterations (epochs) are reached for a convergence training with a mean squared error (MSE) performance goal of 1e-3. The bias value is associated with each node in the intermediate and output layers of a network whose activation is always 1.

In the database used, PCGs durations varied from 1 s to 34 s and down sampled to 1000 Hz with 12107 R peaks and 11554 ends of T waves annotations [13] 270 PCGs (3851.1 s) of 60 normal subjects and 256 PCGs (1696.5 s) of 33 pathological subjects are chosen for training the LRNN. The *T(n)* signal of each PCG is calculated as described in Section II (Fig. 3 (c)), where the positive part of the signum function of the derivative (*dT(n)/dn)* function is taken and used as the training data for the LRNN (Fig. 5 (b)). The labeled signal extracted from the ECG as in Section III, *A* is used as a target of the LRNN. To minimize the calculation time, both the data and targets are down-sampled to 40 Hz.

An example to show robustness of this method can be seen in Fig. 7, where an abnormal PCG with two extra peaks (the first in the diastole of the first heart cycle and the second in the systole of the fourth heart cycle) is considered. From this PCG the variation curve length is calculated *T(n)* (Fig. 7 (b)), then the positive part of the signum function of the derivative (*dT(n)/dn)* function is taken as the input of the LRNN. In this case, the extra-peaks are represented by a positive part of this binary signal in the same way of the FHSs location (Fig. 7 (c)). The output of the LRNN is represented in the form of a stepped signal (Fig. 7 (d)), which indicates the state of each PCG sample. In this figure, it can be seen that the problem of the extra peaks is removed by the LRNN.

The evaluation of the segmentation algorithm is usually based on the calculation of the sensitivity and specificity metrics for the localized FHSs. This is done by determining the True Positive, False Positive, True Negative and False Negative of samples localization.

The evaluation method proposed in [6] and [15] led to the conclusion that the S1 sound was labeled as correctly identified if its start was found to be within 100 ms of the R-peak of the ECG. Likewise, S2 was labeled as correctly identified if found within 100 ms of the end of the T-wave.

In this paper, an evaluation method has been adopted, which is based on the results of segmentation for every time sample at every state. This will enable us to evaluate the partial localization of the S1, S2, systole and diastole states. The state deduced from the ECG will be used as a reference to calculate the sensitivity and specificity for each state (Fig. 6). A True Positive (TP) sample of each state is a sample which is correctly localized by the segmentation method as compared to the ECG labeled state. Otherwise, it is considered as a False Positive (FP). The same process is followed for other states.

$$\text{sensitivity(state)}=100\times\frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

$$specificity(state)=100\times\frac{TN}{FP+TN} \tag{8}$$

The segmentation method with layer recurrent neural network based on the variation of the PCG curve length was tested with a set of 130 PCGs (3191 s) which have a duration that exceeds eight seconds each, and all are from the same database. These are 65 normal PCGs (1612.51 s) and 65 abnormal PCGs (1579.26 s). The evaluation did not take into consideration the first two seconds of any PCG. This time was rejected so as to eliminate the initial samples in order to allow more time for the algorithm to adapt to the signal.

The segmentation evaluation results reported in Table I show 94% specificities and 76% sensitivities performance for all the PCG components (S1, S2, systole and diastole) and almost regardless of its clinical case, being normal or abnormal PCG.

The low performance value of the sensitivity as compared to that of the specificity can be explained by a less appropriate selection of the neural network parameters, such as: the number of neurons in each layer, the number of hidden layers and the transfer function types.

TABLE I
Results of the PCG Segmentation Evaluation (Sensitivity and Specificity (%))

	Sens/ S1	Sens/ S2	Sens/syt	Sens/ diast
Normal	84.35	77.74	80.78	65.18
Abnormal	86.95	74.04	77.50	62.10
	Speci/S1	Speci/S2	Speci/sys	Speci/ diast
Normal	92.47	92.95	93.61	97.05
Abnormal	89.60	94.34	94.06	96.96

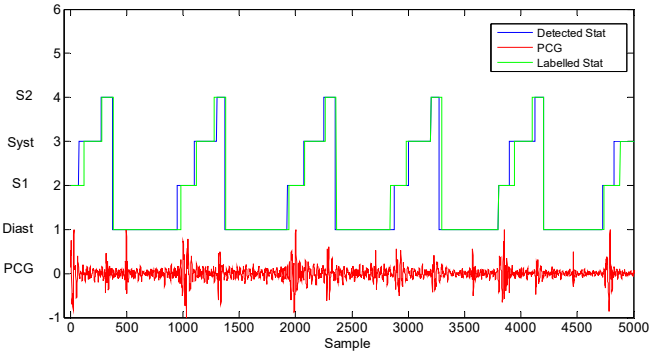

Fig. 6 Results of PCG segmentation with LRNN (blue line) compared with the ECG labeled states (green line)

V. Conclusion

In this paper an automatic segmentation method has been proposed which uses an intermediate signal to localize the Fundamental Heart Sounds S1 and S2 and the systolic and

diastolic components in the PCG signal. This intermediate signal is derived from the variation of the PCG curve length. A Layer Recurrent Neural Network is trained with the positive derivative function of the intermediate signal to achieve a robust segmentation of noisy PCG or with irregular rhythm. Results obtained from the segmentation of a large database show on average a 94% specificity and a 76% sensitivity of the segmentation testing performance.

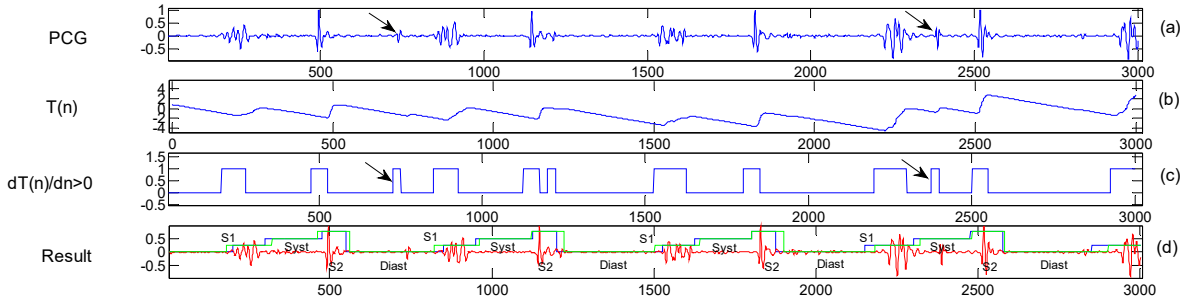

Fig. 7 (a) A normalized abnormal PCG; (b) the T function calculated from the PCG (as explained in Section II); (c) detection of the rising edge of the T function, the positive part of the derivative of T(n) to be used as an input data to the LRNN; (d) results of the output of the LRNN (blue line) with the ECG labeled states (green line)

REFERENCES

[1] World Health Organisation, p. https://www.who.int/cardiovascular_diseases/en/.
[2] S. Choi and Z. Jiang, "Comparison of envelope extraction algorithm for cardiac sound signal segmentation," Expert System, vol. 34, no. 2, pp. 1056, 1069, 2008.
[3] L. Sharma, "Multiscale Analysis of heart sound for segmentation using multiscale Hilbert envelope," in 13 th International Conference on ICT and Knowledge, Bangkok, 2015.
[4] M. Ahfir and I. Kale, "Automatic Synchronization and Segmentation of Heart Sound Signals Based on their Continuous Average Energy," Proc.of The 2nd World Congress on Multimedia and Computer Science, 2014.
[5] M. Z. Belmecheri, M. Ahfir and I. Kale, "Automatic heart sounds segmentation based on the correlation coefficients matrix for similar cardiac cycles identification," Biomedical Signal Processing and Control, vol. 43, pp. 300-310, May 2018.
[6] S. Schmidt, C. Holst-Hansen, C. Graff and E. Tofti, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," Physiol Meas, vol. 31(4), pp. 513-529, 2010.
[7] A. Rowland and A. Sargent, The ECG workbook, M&K Publishing, 2011.
[8] C. Liu, D. Springer and G. Clifford, ""Performance of an open-source heart sound segmentation algorithm on eight independent databases," Institute of Physics and Engineering in Medicin, vol. 38, p. 1730–1745, 2017.
[9] M. Beale, M. Hagan and H. Demuth, "Neural Network ToolboxTM User's Guide," in the Matworks inc, MA, USA, Natick, 2012.
[10] J. Elman, "Finding structure in time," Cognitive Science, vol. 14, p. 179–211, 1990.
[11] "https://www.mathworks.com/help/deeplearning/ref/layrecnet.html," (Online).
[12] Z. Olaofe, "A 5-day wind speed & power forecasts using a layer recurrent neural network (LRNN)," Sustainable Energy Technologies and Assessments, vol. 6, pp. 1-24, June 2014.
[13] D.B. Springer, "heart sound database," https://physionet.org/physiotools/hss, 2016.
[14] L. Resnekov, "Understanding heart sounds and murmurs: with an introduction to lung sounds.," JAMA, vol. 254, no. 1, p. 124–125, 1985.
[15] D. B. Springer, L. Tarassenko and G. D. Clifford, "Logistic Regression-HSMM-based Heart Sound Segmentation," IEEE Transactions on Biomedical Engineering, vol. 63, no. 4, pp. 822-832, April 2016.