2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automatic segmentation for neonatal phonocardiogram

Sergi Gómez-Quintana, Ihor Shelevytsky, Victoriya Shelevytska, Emanuel Popovici, *Senior Member, IEEE,* Andriy Temko[1], *Senior Member, IEEE*

***Abstract*— This work addresses the automatic segmentation of neonatal phonocardiogram (PCG) to be used in the artificial intelligence-assisted diagnosis of abnormal heart sounds. The proposed novel algorithm has a single free parameter – the maximum heart rate. The algorithm is compared with the baseline algorithm, which was developed for adult PCG segmentation. When evaluated on a large clinical dataset of neonatal PCG with a total duration of over 7h, an F1 score of 0.94 is achieved. The main features relevant for the segmentation of neonatal PCG are identified and discussed. The algorithm is able to increase the number of cardiac cycles by a factor of 5 compared to manual segmentation, potentially allowing to improve the performance of heart abnormality detection algorithms.**

I. Introduction

Classification of pathological audio signatures in a phonocardiogram (PCG) recording requires segmentation of the recording into fundamental heart sounds (FHS). FHS usually include the first (S1) and second (S2) heart sounds representing the systolic or diastolic regions. Subsequent feature extraction and classification of pathological audio signatures in these regions allows the identification of various heart conditions [1, 2].

Signal processing methods based on envelopes have been widely used for FHS segmentation [3–5]. Wavelet decomposition has also been employed in [6, 7]. To further improve the accuracy, signal processing methods have been accompanied by machine learning. A deep neural network has been used over a set of mel-frequency cepstral coefficients in [8]. Neural networks have also been used together with features extracted from the discrete wavelet transform [9]. An ensemble of empirical mode decomposition and kurtosis features were used over a significantly larger dataset of thousands of cardiac cycles in [10]. The probabilistic approach based on the hidden semi-Markov model (HSMM) was tested in a dataset of more than 11,000 cardiac cycles [11]. In 2016, a challenge was carried out in Physionet on heart sound classification to address the level of robustness of developed methods to larger datasets, variety of PCG recordings, at times with poor signal quality [12].

While various algorithms have been developed for adult PCG segmentation, very limited work has been done to automate neonatal PCG segmentation. One of the most distinctive characteristics of adult and neonatal PCG is heart rate (HR). Several studies have demonstrated that both HR and maximum HR decline with age. The simplest formula that models the maximum HR in beats per minute (BPM) as a function of age is maximum_HR = 220 – age, which was empirically obtained from data collected in [13]. A subsequent study showed even more accurate formulations using both linear and non-linear approximations [14]. While those studies offer broad models across a wide age range, some studies focused their attention on the edge-subgroups of children and neonates, reporting an HR of 127 BPM (median) in healthy neonates [15] and a maximum HR as high as 222 BPM, but typically 192 BPM [16].

The maximum HR is usually used as an upper bound on the expected HR to better estimate the PCG signal's periodicity and subsequently identify the FHS correctly. Most of the reviewed segmentation algorithms rely on a number of free parameters, and the difference of the maximum HR between adult and neonatal PCG is of significant importance when trying to port an existing adult PCG segmentation algorithm to the neonatal population.

In this work, a novel algorithm for automatic segmentation of neonatal PCG is proposed with a single free parameter – maximum HR. The algorithm is compared with the baseline algorithm [11], which was modified to segment neonatal PCG. The algorithm successfully increases the number of segmented cardiac cycles by a factor of 5 with respect to manual segmentation. The algorithm has been implemented using a similar framework as in [2], which would facilitate subsequent integration into the cloud-based system.

II. Methods

A. Dataset

The dataset used in this study was collected between Sept. 2013 and Sept. 2018 at two hospitals in Ukraine. Informed parental consent was obtained for every participant before study inclusion. The study was approved by local ethics committees. In total, 265 newborns were included in the study, with the gestational ages ranging between 35 and 42 weeks. A variety of diagnoses were present in the recordings, including PDA and various types of CHD, as confirmed by echocardiography.

For each patient, PCG recordings were taken within the first 6 days of life from 5 auscultation areas on the chest using a digital stethoscope recording audio at 44.1kHz and 16bit resolution (Thinklabs ds32a and ThinkLabs One, Centennial, USA). The dataset used in this study consists of 1325 PCG recordings of a total length of 7h 48min, from which only 5904 cardiac cycles are annotated by hand. These amount to 47min

This research was supported by Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No (18/CRT/6223), and the INSIGHT Centre for Data Analytics.

S. Gomez-Quintana, E. Popovici and A. Temko are with Electrical and Electronic Engineering, University College Cork, Ireland, 118227474@umail.ucc.ie (corresponding author), e.popovici@ucc.ie, atemko@ucc.ie.

I. Shelevytsky is with Faculty of Information Technologies, Kryvyi Rih Institute of Economics, Ukraine, sheleviv@gmail.com.

V. Shelevytska is with Faculty of Postgraduate Education, Dnipropetrovsk Medical Academy of Health, Ukraine, shelevika@gmail.com.

TABLE I. DESCRIPTION OF THE NEONATAL PCG DATASET

Total number recordings	**1325**
Recordings per patient	5
Total number of annotated cycles	**5904**
Annotated cycles per recording	Median 5, IQR 4-5
Annotated cycles per patient	Median 23, IQR 21-25
Total hours recorded:	**7h 48min**
Total annotated	47min 53s

which represents approximately 10% of the whole dataset. Only this part can effectively be used for modelling or classification purposes, as was done in [2]. Table I shows the database details.

B. HSMM-based algorithm for adult PCG segmentation

An adult PCG segmentation algorithm [11] has been used as a starting point in this study. The algorithm uses a set of features derived from homomorphic envelogram, Hilbert envelope, wavelet envelopes and the power spectral density of the input PCG signal. Those features are fed into a hidden semi-Markov model (HSMM) that includes a priori statistical information (mean and SD) about the duration to remain at each hidden state (i.e., FHS).

The algorithm in [11] has been modified to make it suitable for faster neonatal PCG signals. In particular, the duration distributions of the FHS and the constraints regarding the maximum HR and the minimum systolic duration were re-tuned on the neonatal dataset.

C. Proposed algorithm

The algorithm is divided into 2 main blocks: segmentation and modelling. The segmentation block performs signal conditioning and splits the audio into sounds (S1 or S2) and silences (systole or diastole), whereas the modelling block aims at resolving the S1/S2 ambiguity.

The flowchart of the segmentation routine can be seen in Figure 1. The main stages are:

- Low-pass filter: the PCG signal is filtered at 1kHz.
- Dynamic-range compression: a compressor is utilized to reduce the dynamic range of the PCG signal amplitudes. For each recording, the 95-percentile on the signal's absolute value is used as an adaptive threshold. The signal is normalized with respect to its maximum amplitude after this process.
- Fast envelope: the signal is convolved using a 2-sided exponential impulse response with an attack/decay time of 15ms. The 15ms are established as 20 times less than the maximum heart rate, 200 BPM.
- Slow envelope: the low-pass filtered version at 3.3Hz of the fast envelope is used to represent the slow envelope. The frequency of 3.3 Hz is equivalent to the maximum heart rate 200 BPM.
- Locate peaks in the slow-envelope as local extrema. The first approximation of the actual HR of the recording is performed using the locations of those peaks and Eq. 1:

$$HR = \text{Median}^{-1}\{\text{s[k]} - \text{s[k} - 2]\}_{1<k<N} \quad (1)$$

  N is the total number of segmented FHS in the recording. HR is estimated from the array of the start times of each FHS s[k] by taking the median of the differences obtained from one out of two consecutive start times (as an S1/S2 alternating sequence is expected).
- Refine the locations of the peaks using the maximum of the fast envelope inside a window of a half duration of a cycle (given by the previously estimated HR).
- Find start and endpoints of the beats (S1 and S2) as the closest point where the fast envelope falls 1/8 from its peak-amplitude (18dB).
- Refine start and endpoints by finding the next zero-crossing point of the original PCG signal (backwards and forward, respectively).

Once the signal is partitioned, the resultant segments need to be classified either as S1 or S2. The modelling routine starts with feature extraction. Sound and silence segments are parametrized using the large set features that capture energy, temporal and frequency information. The same feature set has been utilized for the classification of heart abnormalities in [2]. The top-2 most important features for discrimination between S1 and S2 are the relative time of the silence intervals (RTS) and the tonal deviation from the average central frequency (TDCF) defined in Eq. 2 and Eq. 4:

$$RTS[k] = HR \cdot \text{T}_{silence}[k] \quad (2)$$

where k denotes each segmented silence. The tonal deviation measures the relative difference of the central frequency of the first and second heart sounds with respect to the average central frequency of the recording. The central frequency is computed as:

$$f_c = \frac{f \cdot |X(f)|^2}{\Sigma |X(f)|^2} \quad (3)$$

where f is the array of frequencies from 5Hz to 200Hz and $X(f)$ the discrete Fourier transform of the segmented sound (S1 or S2). After the central frequency is obtained, the tonal deviation for each segmented sound, k, in the recording is obtained as:

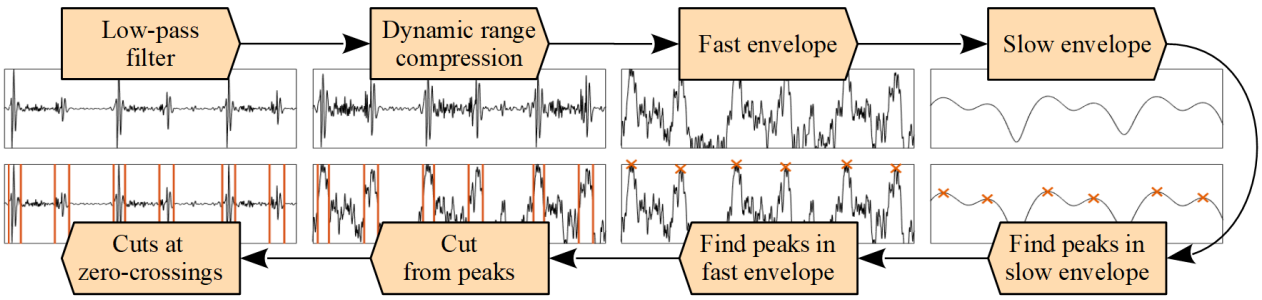

Figure 1. The flowchart of the segmentation algorithm.

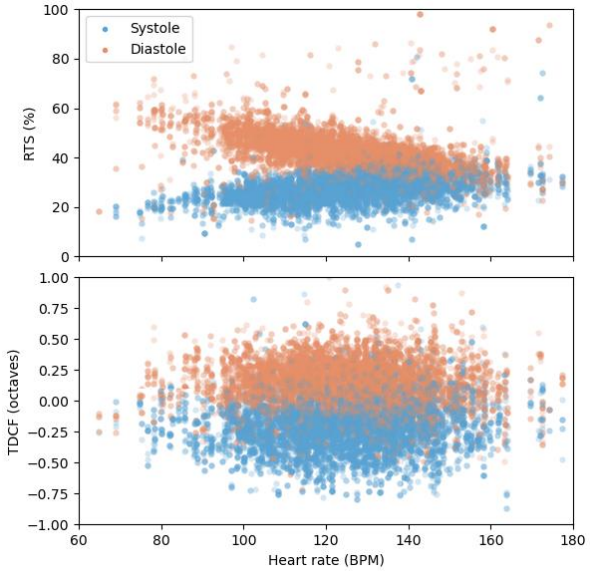

Figure 2. Two most important features for Systole-Diastole classification as a function of HR: RTS (top), TDCF (bottom). Best seen in color.

$$TDCF[k] = \log_2 \frac{f_c[k]}{mean(f_c[k])_{0 \le k \le N-1}} \quad (4)$$

The logarithm base 2 is used to obtain the relative difference in octaves – a tonal deviation of 1 octave implies that the central frequency is twice the average central frequency.

A gradient boosting decision tree (XGBoost) classification algorithm is trained and tested using subject-independent 10-fold cross-validation (CV), as described in [2]. The models were constructed using the following settings: objective = binary: logistic, eval_metric = auc, eta = 0.03 (learning rate). Other hyperparameters such as the number of branch levels for each decision tree (max_depth), a ratio of randomly selected data rows or samples (subsample), a ratio of randomly selected data columns or features (colsample_bytree), and the number of decision trees used by the model (tree_num) are tuned using the nested CV routine.

III. Results

Table II compares the developed algorithm performance for neonatal PCG segmentation with that of the HSMM-based algorithm [11]. The performance of the latter is obtained with and without parameter tuning on neonatal PCG. The proposed algorithm achieves the best score for all considered metrics.

Table III outlines the number and the values of tunable hyperparameters of the compared algorithms both for neonatal and adult PCG. It can be seen that, while the HSMM algorithm hyperparameters are estimated from the dataset itself, the proposed algorithm has only a single free parameter and this can be obtained as function of the patient's age, which is well reported in the related literature [13–16].

TABLE II. The comparison of Algorithms for Neonatal PCG Segmentation

Algorithm	Se	P+	Acc	F1	F1(S1)	F1(S2)
HSMM (orig)	0.61	0.75	0.60	0.65	0.77	0.53
HSMM (mod)	0.91	0.93	0.90	0.91	0.93	0.89
This work	0.93	0.95	0.92	0.94	0.94	0.93

The reported metrics to assess the accuracy of adult PCG segmentation are: recall, also known as sensitivity (Se), precision, also known as positive predictive value (P+), accuracy (Acc) and F1 score over precision and recall (F1). The latter is also reported separately for the first and second heart sounds as F1 (S1) and F1 (S2), respectively.

Figure 2 shows the scatter plot of Systole and Diastole classes for the top-2 features against the HR, RTS and TDCF. Diastolic RTS tend to be larger than systolic RTS at low HR. On the range between 130-160 BPM, both systolic and diastolic RTS converge to a similar relative duration, increasing overlap between the two classes. The TDCF difference between S1 and S2 is half of an octave, on average, with no dependence on HR.

Figure 3 illustrates an example of 5 cycles of neonatal PCG segmented by hand, the proposed algorithm and the two variants of the HSMM algorithms.

IV. Discussion

Automated segmentation of neonatal PCG to its FHS can improve the accuracy of the heart abnormality detection models through the generation of more training data. The usage of this algorithm can potentially expand the current dataset from 47min 53s minutes to 262 min 33s which represents an increase in the number of cardiac cycle instances by a factor of 5. Data-hungry deep learning approaches can be considered for the classification of neonatal heart sounds [17]. More importantly, automated methods to segment neonatal PCG can also remove the burden on the need for a healthcare professional to segment each cycle, which is a key step for the AI-based decision support systems. It has been shown in [2] that approximately half of the data used in this study contained noises external to the heart sounds, making the task of segmentation very challenging, both for the ML and the healthcare professional. Those external noises included mainly baby crying, people speaking, or movement artefacts (e.g. skin scratching).

It can be seen from Figure 3, the HSMM-based algorithm, HSMM (orig), tuned on adult PCG and tested on neonatal PCG estimates a wrong HR as half of the actual HR. As a result, it often fails to detect the S2 segment correctly and only detects one out of the two S1 in the shown example. The algorithm has considerably lower F1 scores, especially when it comes to identifying the second heart sound (S2).

When HSMM-based algorithms parameters are properly adjusted, HSMM (mod), a boost in performance can be observed from Table II. Figure 3 also shows a reasonably accurate segmentation obtained with HSMM.

TABLE III. Hyperparameters and their values used for algorithm performance assessment.

HSMM [11]		
Hyperparameter	**Adult PCG**	**Neonatal PCG**
S1 duration (mean)	122ms	78ms
S1 duration (SD)	22ms	20ms
S2 duration (mean)	92ms	51ms
S2 duration (SD)	22ms	15ms
Min. systolic duration	200ms	100ms
Maximum HR	120 BPM	200 BPM
This work		
Hyperparameter	**Adult PCG**	**Neonatal PCG**
Maximum HR	120 BPM	200 BPM

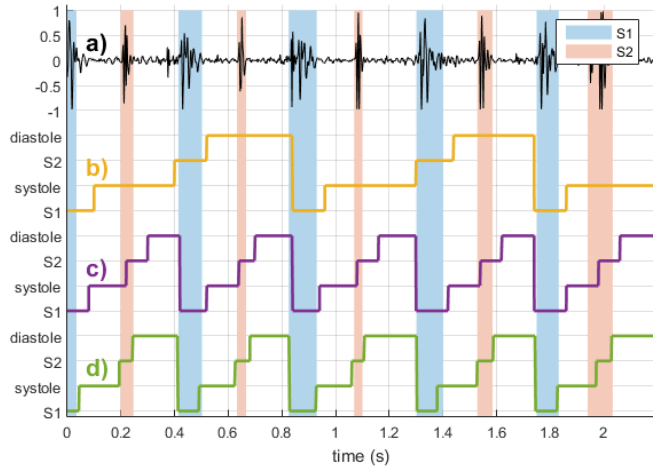

Figure 3. An example of segmentation applied on a 5-cycle PCG recording. Blue and red areas indicate the ground truth for the segmentation of S1 and S2 sounds, respectively. From top to bottom, the graph shows the following traces: PCG waveform (a); the sequence of states obtained with the original (b) and adapted HSMM-based algorithms (c); the sequence of states obtained with the proposed algorithm (d). Best seen in color.

However, the HSMM-based algorithm required several hyperparameters to be adjusted to fit the neonatal PCG characteristics. First, a priori knowledge of the durations of S1 and S2, which are not HR dependent, are required. The algorithm uses this information to estimate the time to remain in each state (S1, systole, S2, diastole). The durations obtained from the adult dataset [11, 18] were 122ms for S1 and 92ms for S2, both within a standard deviation of 22ms. When it comes to neonatal heart sounds, those durations are significantly smaller, and the average values are estimated to be 78ms for S1 and 51ms for S2, with 20ms and 15ms standard deviations, respectively. Moreover, the minimum systolic duration was assumed to be 200ms for adult PCG in the original algorithm. This had to be reduced to 100ms in the modified version of the HSMM-based algorithm.

In contrast, the proposed algorithm requires no prior knowledge of the S1 and S2 durations. The algorithm only requires the maximum HR specified. Figure 3 shows that the developed algorithm better captures the onset and offset of S1 and S2 segments, which is reflected in the highest F1 scores both for S1 and S2, in Table II.

The developed algorithm was also tested on the dataset of adult PCG by setting the maximum HR to 120 BPM. An F1 score of 0.91 was achieved as compared to an F1 score of 0.96, which was reported with the HSMM-based solution. With a single tunable hyperparameter, the algorithm works comparably well in a variety of scenarios.

The systolic duration is known to be shorter than the diastolic duration in most cases. Some heart sound segmentation methods are entirely based on that assumption [6, 9]. Several studies have shown, however, that for fast HR, the opposite can occur [19, 20]. Our findings are consistent with these reports, as can be seen in Figure 2 for RTS. For HR between 130-160 BPM, the systolic time can be of at least the same duration as the diastolic time. The proposed method incorporates additional features that are not HR-dependent, such as the TDCF, allowing more accurate differentiation, especially in the neonatal PCG, where the HR is faster when compared to adult PCG.

V. Conclusions

This paper presents a novel approach to segment neonatal PCG. The method compares favourably with existing algorithms in terms of the performance and the number of parameters to tune. The segmented PCG can increase the amount of annotated cardiac cycles 5-fold to improve heart abnormality detection algorithms' accuracy. The developed method represents a promising step towards fully-automated AI-based decision support for PCG-based detection of neonatal heart abnormalities.

References

[1] C. Potes et al., "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," *in Proc. Comput. Cardiol.*, pp. 621–624, 2016.

[2] S. Gomez-Quintana et al., "A framework for AI-assisted detection of patent ductus arteriosus from neonatal phonocardiogram," *Healthcare*, 9(2), p. 169, 2021.

[3] A. Iwata et al., "Algorithm for detecting the first and the second heart sounds by spectral tracking," *Med. Biol. Eng. Comput.,* 18(1), pp. 19–26, 1980.

[4] H. Naseri, M. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, 41(2), pp. 279–292, 2013.

[5] A. Moukadem et al., "A robust heart sounds segmentation module based on S-transform," *Biomed. Signal Proces.*, 8(3), pp. 273–281, 2013.

[6] V. Varghees, K. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sens. J.*, 17(12), pp. 3861–3872, 2017.

[7] S. Patidar, R. Pachori, "Segmentation of cardiac sound signals by removing murmurs using constrained tunable-Q wavelet transform," *Biomed. Signal Proces.*, 8(6), pp. 559–567, 2013.

[8] T. Chen et al., "S1 and S2 heart sound recognition using deep neural networks," *IEEE Trans. Biomed.*, 64(2), pp. 372–380, 2017.

[9] S. Yuenyong et al., "A framework for automatic heart sound analysis without segmentation," *Biomed. Eng. Online*, 10(1), pp. 1–23, 2011.

[10] C. Papadaniil, L. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health*, 18(4), pp. 1138–1152, 2014.

[11] D. Springer et al., "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed.*, 63(4), pp. 822–832, 2016.

[12] C. Liu et al., "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, 37(12), pp. 2181–2213, 2016.

[13] S. Fox, J. Naughton, "Physical activity and the prevention of coronary heart disease," *Prev. Med.*, 1(1–2), pp. 92–120, 1972.

[14] R. Gellish et al., "Longitudinal modeling of the relationship between age and maximal heart rate," *Med. Sci. Sports & Exerc.*, 39(5), pp. 822–829, 2007.

[15] S. Fleming et al., "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies," *The Lancet*, 377(9770), pp. 1011–1018, 2011.

[16] T. Montague et al., "The spectrum of cardiac rate and rhythm in normal newborns," *Pediatr. Cardiol.,* 2(1), pp. 33–38, 1982.

[17] Y. Lecun et al., "Deep learning," *Nature*, 521, pp. 436–444, 2015.

[18] S. Schmidt et al., "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, 31(4), pp. 513–529, 2010.

[19] W. Cui et al., "Systolic and diastolic time intervals measured from Doppler tissue imaging: normal values and z-score tables, and effects of age, Heart Rate, and Body Surface Area," *J. Am. Soc. Echocardiogr.*, 21(4), pp. 361–370, 2008.

[20] M. Friedberg, N. Silverman, "The systolic to diastolic duration ratio in children with heart failure secondary to restrictive cardiomyopathy," *J. Am. Soc. Echocardiogr.*, 19(11), pp. 1326–1331, 2006.