Discrimination of Normal and Abnormal Heart Sounds Using Probability Assessment

Filip Plesinger, Juraj Jurco, Pavel Jurak, Josef Halamek

ISI of the CAS, Brno, Czech Republic

Abstract

Aims: According to the "2016 Physionet/CinC Challenge", we propose an automated method identifying normal or abnormal phonocardiogram recordings.

Method: Invalid data segments are detected (saturation, blank and noise tests). The record is transformed into amplitude envelopes in five frequency bands. Systole duration and RR estimations are computed; 15-90 Hz amplitude envelope and systole/RR estimations are used for detection of the first and second heart sound (S1 and S2). Features from accumulated areas surrounding S1 and S2 as well as features from the whole recordings were extracted and used for training. During the training process, we collected probability and weight values of each feature in multiple ranges. For feature selection and optimization tasks, we developed C# application PROBAfind, able to generate the resultant Matlab code.

Results: The method was trained with 3153 Physionet Challenge recordings (length 8-60 seconds; 6 databases). The results of the training set show the sensitivity, specificity and score of 0.93, 0.97 and 0.95, respectively. The method was evaluated on a hidden Challenge dataset with sensitivity and specificity of 0.77 and 0.91, respectively. These results led to an overall score of 0.84.

1. Introduction

This paper describes a method for automated discrimination of normal and abnormal heart sounds, according to Physionet Challenge 2016 [1]. The presented method introduces a probability assessment from multiple features. The used features are mostly derived from the accumulated systole and diastole area. Moreover, some features are independent of heart sounds detection.

2. Method

The workflow of the method is presented in Fig. 1. It consists of the generation of amplitude envelopes in 5 frequency bands, detection of invalid areas, detection of heart sounds S1 and S2, extracting features, limit test and computing and comparing the resultant probability.
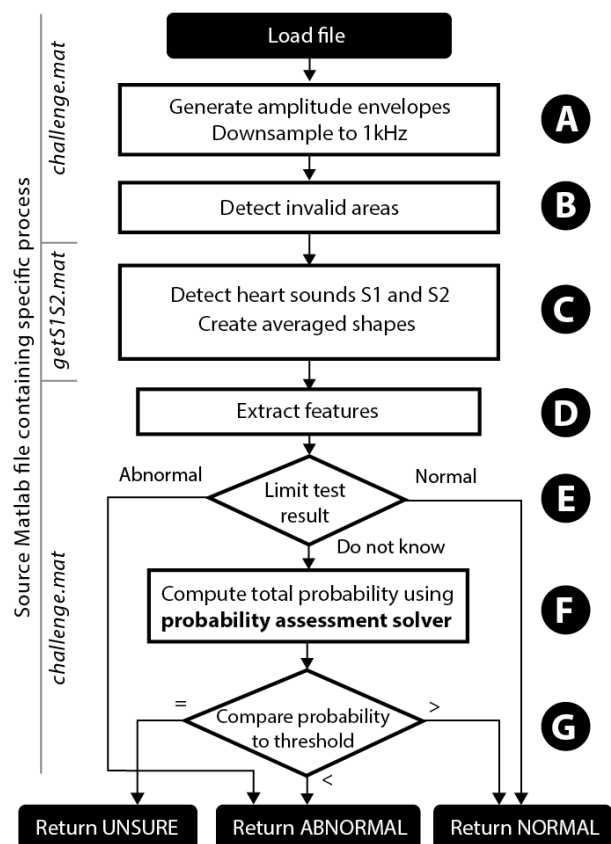

Figure 1. Flowchart of presented method

2.1. Pre-processing

Five amplitude envelopes are generated from the input wave file (sampling at 2 kHz) using Hilbert and Fourier transformations. The generated envelopes use the bands 15-90 Hz (LF), 55-150 Hz (MF), 100-250 Hz (HF), 200-450 Hz (SF) and 400-800 Hz (UF). All generated envelopes and source signal are down-sampled to 1 kHz (Fig. 1A). The LF envelope is used in S1 and S2

detection; the EF and LF ratio is used for invalid area detection. The LF, MF, HF, UF and SF envelopes (Fig. 2) are used in the limits check and probability assessment solver.
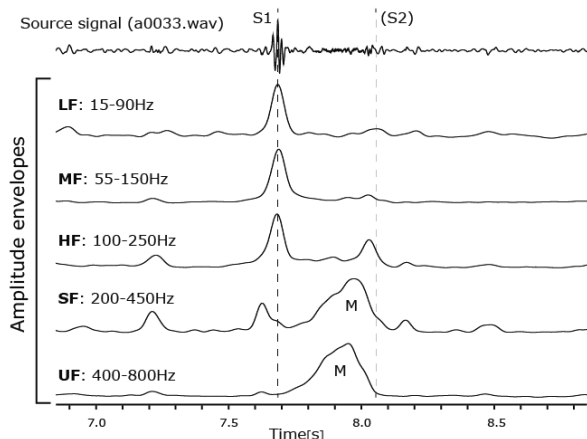

Figure 2. Abnormal heart sounds (record a0033.wav) and computed amplitude envelopes (smoothed with LP filter). S1 and S2 are the first and the second heart sound, respectively. M is audible murmur, visible in the two highest frequency bands SF and UF. The image was rendered using SignalPlant software [2].

2.2   Detection of invalid areas

Invalid areas are detected in a 1 s moving window with a 0.5 s step (Fig. 1B). A histogram (10 bins) from the raw signal is made. If the histogram leads to a non-usual signal distribution (based on [3]) or the sum of the HF envelope is higher than the sum of the LF envelope, the block is marked as "invalid".

2.3.   S1 and S2 heart sound detection

Prior to the heart sound detection, an estimation of the systole duration and RR interval is made. Systole duration estimation S_{DE} (Fig. 3A) uses the LF amplitude envelope. Peaks are detected (threshold 74th percentile, minimal peak distance 175 ms) and a histogram of their distances is built. Due to the low variability of the systole duration, the most common bin in the histogram should lead to systole duration. RR estimation (RR_E) is based on autocorrelation of the LF envelope (Fig. 3B). In the resultant autocorrelation function peaks are found (threshold 80th percentile, min. peak distance 300 ms). Peak differences are collected (excluding non-physiological RR distances) and the RR interval is computed as a median of the peak distances. In case of fewer than 3 distances, the estimated RR interval is computed as 2.8 x $S_{DE.}$ The RR to systole ratio is compared (Fig. 3C) to the limits (> 2.1; < 4.5). If it does

not pass, the S_{DE} or RR_E is corrected. The heart sound extraction block (Fig. 3D) searches in peaks derived from the LF envelope and finds S1 and S2 pair locations. If the number of resultant pairs is lower than 2, the process is aborted (Fig. 3E).
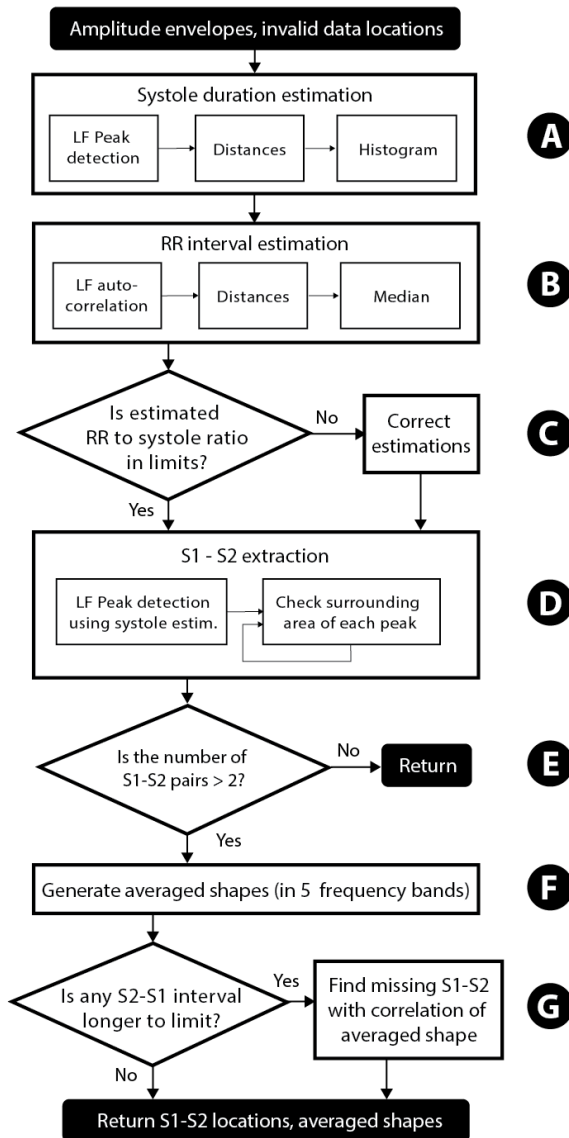

Figure 3. Flowchart of proposed heart sound detector.

Averaged shapes of systole and diastole area are generated in each of the used frequency bands (Fig. 3F).

These averaged shapes are intended for feature collection as well as for missing S1/S2 detection check. If any S2-S1 interval (i.e. diastole) is longer than 1.8 s (Fig. 3G), the LF average shape of the S1-S2 area is used to find missing heart sounds with correlation (threshold 0.8).

2.4. Investigated features

We extracted 215 features for further investigation. These features were statistical properties of segments of averaged heart sound envelopes in several time segments (before S1, at S1, after S1, before S2, at S2, after S2 and diastole) in five frequency ranges. The ratios between those properties were also used.

Moreover, we generated features independent of detection of S1 and S2. We tested the skewness and kurtosis of all the generated envelopes. We also designed a transformation reacting to continuous blocks of a similar sound:

$$CS_i = Log\left(Corr(FFT_i, FFT_{i-step}) \times \sum FFT_i\right)$$

where FFT_i is the absolute value of the Fast Fourier transform (window 128 ms) following a sample with index i. This transformation is able to reveal murmurs, as shown in Fig. 4.
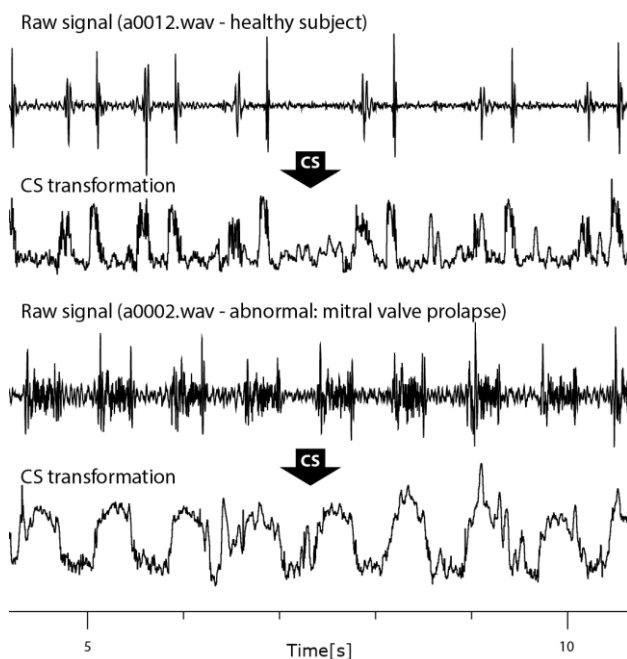

Figure 4. CS transformation of a signals containing normal and abnormal heart sounds. It positively reacts to continuous sounds of a similar frequency image, revealing murmurs.

2.5. Limit test – preliminary classification

Some of the extracted features were used for the raw classification of normal (abnormal) sounds (Fig. 1E). Feature thresholds were set to separate only normal (resp. abnormal) records from the rest. If this limit test points to normality (abnormality), further processing is skipped.

Table 1 shows the number of futures used for the limit test and its ability to reveal normal and abnormal sounds for each tested dataset.

Table 1. Count of features used for the preliminary classification of normal (N_{PN}) and abnormal (N_{PA}) sounds in each dataset. No false negative detection was allowed during training. N (A) is the relative number of normal (abnormal) sounds in the dataset. All (1) features were independent on heart sounds detection, while All (2) used them. Datasets A and F were excluded due to their absence in test set; C was excluded because it can be divided using only one parameter.

Dataset	N_{PN}	% of N	N_{PA}	% of A
All (1)	2	40%	1	17%
All (2)	4	33%	7	16%
B	4	16%	3	16%
D	2	54%	4	25%
E	3	96%	5	66%

2.6. Training - probability assessment solver

Because no single feature (except dataset C) could be used for a reliable division into normal and abnormal sounds, multiple features were used for each dataset. Multiple features usually lead to machine learning methods such as neural networks [4] or supported vector machines [5]. Contrary to these standard approaches, we decided to design a new method using the probability of a normal/abnormal sound in single sections in each feature.
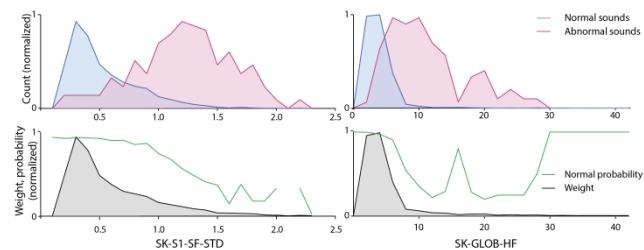

Figure 5. Example of two features (dataset E). Left – standard deviation of skewness of averaged S1 surroundings in the SF frequency band, right – skewness of the record HF amplitude envelope. Blue - occurrence of normal sounds, red - occurrence of abnormal sounds, black - weight, green – normal sound probability.

For each feature, two histograms (20 bins) were built for the normal/abnormal values (Fig. 5). The weight of each histogram bin was computed as the ratio between normal and abnormal occurrences in the bin to the total number of subjects. The resultant probability is transformed with the sigmoid function and added to the probability from the other parameters (Fig. 1F). The summed probability is compared to the manually selected

threshold (Fig. 1G) leading to a final statement about heart-sounds normality. For the inspection and automated feature selection process, we built PROBAfind software (http://medisig.com/probafind). It also allows interactive setting of the sensitivity/specificity ratio and generates a Matlab code for challenge entry.

2.7. Probability assessment - selected features

Due to the different behaviour of the same features in different datasets (containing same pathology, as in Fig. 6) we decided to build separate feature set for each dataset. For test-only datasets G and I, general feature set was built. Features were selected semi-automatically using PROBAfind software. Table 2 shows the number of used features and the strongest feature for the training datasets. Dataset C (containing mitral regurgitation and mitral stenosis) was not subjected to the probability assessment; it can be divided using just one parameter.

Table 2. Selected features for training datasets. N shows the total number of selected features for each dataset. "Other" is general dataset in case that the dataset is not known or cannot be identified. Dataset C does not use probability assessment, it uses simple comparison instead.

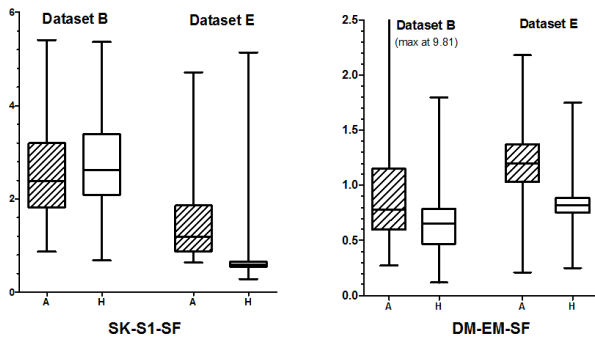| Set	N	The strongest feature
A	13	Diastole value in UF to LF ratio (median)
B	11	Ratio between value in S1 to after-S1 area in accumulated LF (median)
C*	1	Skewness of accumulated S2 in HF
D	2	Ratio between value in S1 to after S1 area in accumulated LF
E	4	Skewness of accumulated S1 in UF
F	7	Mean of accumulated area after S2 in LF
Other	13	Skewness of accumulated S1 in UF

Figure 6. Feature heterogeneity in datasets B and E with coronary artery disease (CDA). Presented features (SK-S1-SF - skewness of accumulated S1 area in SF band; DM-EM-SF - ratio between S2 and the preceding area in the SF band) separate healthy subjects (H) from CDA patients (A) in dataset E. The same parameters are less usable in dataset B.

3. Results and conclusion

The training and testing results are presented in Table 3. Testing was carried out on a hidden test set. The presented method consumed 5.11% (average per record) of instructions allowed.

Table 3. Training and testing results. N – number of records, Se – sensitivity, Sp – specificity.

Dataset	N	Se	Sp	Score
Training set	3153	0.93	0.97	0.95
Test set (hidden)	500	0.77	0.91	0.84

An approach for the automated detection of normal/abnormal heart sounds using a probability assessment was introduced. We also proposed features independent of heart sound detection, which may be suitable for hardware implementation.

Acknowledgements

References

[1] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. W. Johnson, Z. Syed, S. E. Schmidt, C. D. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, M. R. Samieinasab, R. Sameni, R. G. Mark, and G. D. Clifford, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 9, 2016.

[2] F. Plesinger, J. Jurco, J. Halamek, and P. Jurak, "SignalPlant: an open signal processing software platform," *Physiol. Meas.*, vol. 37, no. 7, pp. N38–N48, 2016.

[3] F. Plesinger, P. Klimes, J. Halamek, and P. Jurak, "Taming of the monitors: reducing false alarms in intensive care units," *Physiol. Meas.*, vol. 37, no. 8, p. 1313, 2016.

[4] J. P. De Vos and M. M. Blanckenberg, "Automated pediatric cardiac auscultation," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 244–252, 2007.

[5] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support Vectors Machine-based identification of heart valve diseases using heart sounds," *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 47–61, 2009.

Address for correspondence:
Filip Plesinger, ISI of the CAS
Kralovopolska 147,
Brno 612 64, Czech Republic
E-mail address: fplesinger@isibrno.cz