Recognition of Abnormalities in Phonocardiograms for Computer-Assisted Diagnosis of Heart Failures

Simon Hofmann, Volker Groß, Andreas Dominik

THM University of Applied Sciences, KITE Kompetenzzentrum für Informationstechnologie, Gießen, Germany

Abstract

We present algorithms to distinguish between healthy and diseased condition of the heart, based on the analysis of phonocardiograms. The software tries to mimic the decision-making process of a cardiologist by identifying heart beats (S1 and S2), finding extra sounds and murmurs while ignoring all kinds of artefacts and noise.

Two different solutions have been submitted to the PhysioNet Challenge 2016: The entry for phase I aims to reconstruct the signal of an ideal heartbeat by calculating the median of an overlay of all beats of a recording. An LVQ-classifier, trained with the ideal beat of 3240 PCGs of the challenge training set, achieved a specificity of 0.85 and a sensitivity of 0.40, resulting in a total score of 0.63.

Our entry for the official phase of the challenge searches for abnormalities in every single beat of a PCG. The results display a sensitivity of 0.91, a specificity of 0.29, and a total score of 0.60.

1. Introduction

Examination of heart sounds by use of a binaural stethoscope is a common part of almost every medical check-up. Bedside or long term monitoring of heart disease patients, however, is still mainly focussed on continuous ECG, although valve diseases, which show an increasing prevalence in ageing societies, can be detected inexpensively by evaluating phonocardiograms (PCGs) [1].

In contrast to electrocardiograms (ECGs), for which standardised protocols ensure constant quality of the signals, the signal quality of a (PCG) is often variable. Recordings are frequently interspersed or even dominated by background noise, respiratory sounds or other disturbances, which have similar spectrum and energy properties as heart sounds and murmurs [2].

2. Algorithm

Phase I entry: A first approach to read parameters from noisy and low quality recordings is to uncover the signal of an undisturbed ("idealised") heartbeat for each PCG.

The first step of the algorithm is the detection of heartbeats. For further analysis only the intensity of the signal, estimated as Hilbert amplitude envelope, is considered. Windows comprising single beats are extracted, superimposed and averaged. The final idealised beat is cleaned from random artefacts as well as from background noise and short-term disturbances.

Descriptors derived from this single beat by binning are used for training models by application of learning vector quantisation (LVQ) [3] and multilayer perceptrons (MLP) [4].

Official phase entry: A second algorithm analyses the individual heartbeats of a PCG recording regarding presence of abnormalities, such as extra sounds, murmurs or split sounds.

The identification is based on frequency and time binning over each heart cycle and subsequent searching for relations between signal components [5].

Identified abnormalities of each kind are summed up and result in a feature vector with 8 values (S3, S4, systolic extra sound (SeS), diastolic extra sound (DeS), systolic murmur (SyM), diastolic murmur (DiM), S1 split (S1-split) and S2 split (S2-split)) which describe the level of abnormalities of each PCG. Values for S3, DeS and the number of split sounds are normalised by the number of beats found. Values for S4, SeS and the murmurs are not normalised as even rare events are considered to indicate pathological status.

A classifier is trained to build a predictive models that discriminates between *Normal* and *Abnormal*.

2.1. Dataset

Training of the predictors is performed with the dataset of the PhysioNet Challenge 2016 that comprises 3240 PCGs of healthy subjects and patients (2575 annotated as *Normal*, 665 as *Abormal*). The dataset unites PCGs from different sources and provides a heterogeneous training base, comparable to what the software will encounter in

clinical practice [6].

2.2. Frequency analysis

The frequency bands are extracted by a windowed Fourier transformation with a Hanning window of length 128ms by using the R packages `seewave` and `fftw` without zero-padding [7, 8]. In order to make the analysis independent from sampling rate and to speed up subsequent computational steps, the resulting time-dependent frequency spectrum is resampled to a rate of 100 Hz.

Amplitudes of frequency bands are derived from the spectrogram by integrating over 10 ms bins and the frequency ranges given in table 1.

Table 1. Frequency bands for beat and feature recognition.

Name	Frequencies [Hz]	Name	Frequencies [Hz]
low	20–50	hf3	450–550
beat	20–250	hf4	550–650
high	50–150	hf5	650–750
hf1	250–350	hf6	750–850
hf2	350–450	hf7	850–950

2.3. Heart beat detection

For the phase I entry beat detection is performed based on the envelope, for the official phase entry based on the frequency band *beat* in several steps: Firstly, a baseline correction is applied by subtracting a spline, fitted through the minima of the band. Next, the band is normalised to the 95%-percentile. Then, raw beats are extracted as local maxima of 100 ms windows after verification that there is no background noise (i.e. artefact) in this region.

Beats are extracted from the resulting list of raw signals in two steps: As the S1-S2-distance is considered to be the most conserved value in the raw series of sounds, S1-S2 is defined as the maximum of the density function of differences of consecutive raw sounds in a range of 180–400 ms. The S1-S1-distance is estimated from the list of S1-S2-pairs as maximum of the density of the differences of S1 values in a range between S1-S2 and 1.5 s.

Finally a raw S1-S2-pair is only accepted as a beat, if a second pair is present before or after in a distance of approximately S1-S1.

In addition, a minimum of 0.6 beats per second is demanded. Both bands *beats* and *low* are analysed and the filter limits are iteratively weakening, if necessary.

2.4. Artefact elimination

For artefact detection, the energy in the bands *hf2, hf3, hf4, hf5, hf6, hf4* is determined for a time window (of 60 ms) around a raw beat or feature. If the energy of a band is 1.2 times higher than the global average of the respective band, it is marked as active. If at least four bands are active for more than 50% of the time window, the sound is treated as an artefact.

During raw beat detection all artefacts are checked for equal distribution, as equal distributed artefacts may indicate long term high frequency murmur.

2.5. Feature detection

The data preparation (frequency analysis, beat detection, artefact elimination) delivers a dataset comprising of the frequency bands and a list of heartbeats. This dataset allows for analysing each single beat for presence or absence of features linked with disease:

S3 extra sounds are diastolic sounds, observable in the PCG after S2. The filter detects S3 by searching for peaks in the band *beat* from 110–210 ms after S2.

S4 extra sounds are disatolic sounds hearable before S1. S4 are searched as peaks in the band *beat* at 220–80 ms before S1.

Other **diastolic extra sounds (DeS)** are defined as peaks in the band *beat* in a range of 50–110 ms after S2 (opening sound from mitral or tricuspid valve).

Systolic extra sounds (SeS) are defined as peaks in the band *beat* between S1 and S2 with a minimum distance of 75 ms to S1 and S2.

Systolic (SyM) is identified, if a minimum of three of the bands *hf1, hf2, hf3, hf4* show an energy of 10% of the average of S1 and S2 in band *beat* for at least 20% of the time between S1 and S2. Is the duration less then 20%, the median energy of band *high* is computed before, after and between the heart cycle and compared with one another.

Diastolic murmur (DiM) is searched for in a time range from 70 ms after S2 to 20 ms before the next S1. DiM is identified, if a minimum of three of the bands *hf1, hf2, hf3, hf4* show an energy of 10% of the average level of S1 and S2 in band *beat* for at least 10% of the time observed.

Splitted heart sounds (S1-split, S2-split) are detected, if in a range of $\pm$ 100 ms around S1 or S2 two maxima are detected.

2.6. Classification

After feature detection, each PGC is characterised by an 8-dimensional vector consisting of one score per feature. The vectors are z-score normalised and a LVQ model with 100 codebook vectors is trained for 5000 steps with the lvq1- and for additional 5000 steps with the lvq2-algorithm [3, 9]. Prediction is performed with a k-nearest neighbour classifier that allows for selecting the number of neighbours k and a distance threshold. During development the model is tested applying 33% cross-validation.

A multilayer perceptron comprising 2 hidden layers with 12 and 8 neurons is trained as additional classifier [4].

3.   Results

The **phase I entry** has been applied to the 3126 PCGs of datasets a–e of the unofficial phase of the challenge. Envelope-based beat detection was successful for 3035 PCGs. No beats were found in 91 PCGs. Figure 1 shows an example for the reconstructed ideal beat of an *abnormal* PCG. The blue bins visualise 19 descriptive values used for classification.
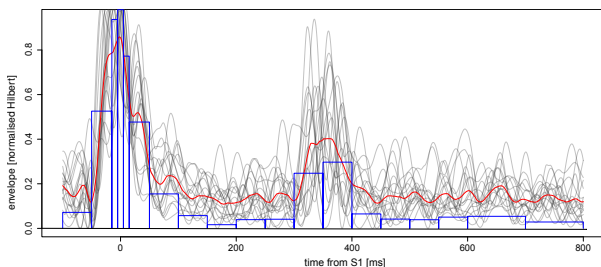

Figure 1. Superimposed heart beats of sample a0058: Although the recording shows differences from beat to beat, an idealised signal can be reconstructed (red curve). The blue boxes indicate the bins used for classification.

Cross validated results on the challenge training set, using LVQ and a 1-nn predictor give a selectivity of 0.65, a specificity of 0.76 and a score of 0.70. The same model achieves a score of 0.63 on the hidden validation set of the challenge. Comparable results are achieved with the neural net: selectivity: 0.68, specificity: of 0.74 and score: 0.71.

Beat detection of the 3240 PCG recordings of the challenge datasets a–f have been performed with the **official phase entry**. A minimum of 6 beats was found for 3035 PCGs. Cross validated results on the training set display a specificity of 0.29, a sensitivity of 0.91 and a score of 0.60. The score for the hidden validation set is identical to the cross-validated result (sensitivity 0.29, specificity 0.91, overall 0.60).

As an example, identified abnormalities for recording b0106 (annotated as *Abnormal*, patient suffering from CAD) are listed in table 2. Figure 2 displays raw waveform, spectrogram and amplitudes of the frequency bands. The beats before 0.5 s and after 7.0 s are not selected by the algorithm, because of background noise. S1 splits at 0.65 s and 3.05 s as well as S2 splits at 5.8 s and 6.4 s are detected in the band *beat*. An S3 extra sound is recognised at 6.6 s in the bands *beat* and *low*. A diastolic murmur is detected after S2 at 6.93 s.

Table 2. Detected features for an example PCG (b0106 of the challenge dataset annotated as *Abnormal*). S1-S2 is estimated as 0.268 s, S1-S1 as 0.600 s.

S1 [s]	S2 [s]	Features
0.655	0.915	S1-split
3.045	3.315	S1-split
5.485	5.755	S2-split
6.085	6.355	S3, S2-split
6.665	6.925	DiM

4.   Discussion

The basic idea behind our phase I entry was the hypothesis that heart diseases lead to typical distortions of the heartbeats in PCGs, which are present in all beats but not always recognisable because of noise and other sounds. We could not verify this hypothesis. In contrast, it turns out that for most patients the abnormalities appear only infrequently. Consequently, abnormalities are averaged out during reconstruction of idealised beats and as a result a majority of the idealised *Abnormal* beats still look *Normal*.

Knowing this, the software should not be able to detect any of the *Abnormal* recordings. Most probably the (limited) predictive power of the entry results from bias in the dataset, like the high correlation between patient age annotation (almost all *Normals* are young persons; almost all *Abnormals* are elderly patients). Correlation between age and heart rate variability (HRV) is well known and HRV is registered in the idealised beats as width of the peaks. We therefore assume that the software predicts the age of study subjects rather than their diseases.

As a consequence we have built the official phase entry differently with the aim to detect abnormal features in every single beat. The resulting tool reliably detects abnormalities (i.e. extra sounds, splits and murmurs) with a high sensitivity (>90% of *Abnormal* PCGs are recognised). However, abnormal features are detected in hundreds of *Normal* recordings, too. Detailed inspection did not reveal significant differences between true disease and the false positives, neither in frequency of features nor in strength of the signals.

The only interpretation of this is that many of the PGCs marked as *Normal*, display evidence for heart disease, which however is not audible during a manual examination with a stethoscope or is an innocent or benign murmur. Further investigation is necessary to decide, if these subjects need to be re-classified because of a present (but so far unrecognised) heart disease.

Of course, more information, such as knowledge of the location of auscultation, position and age of the patient, is essential for reliable diagnosis.
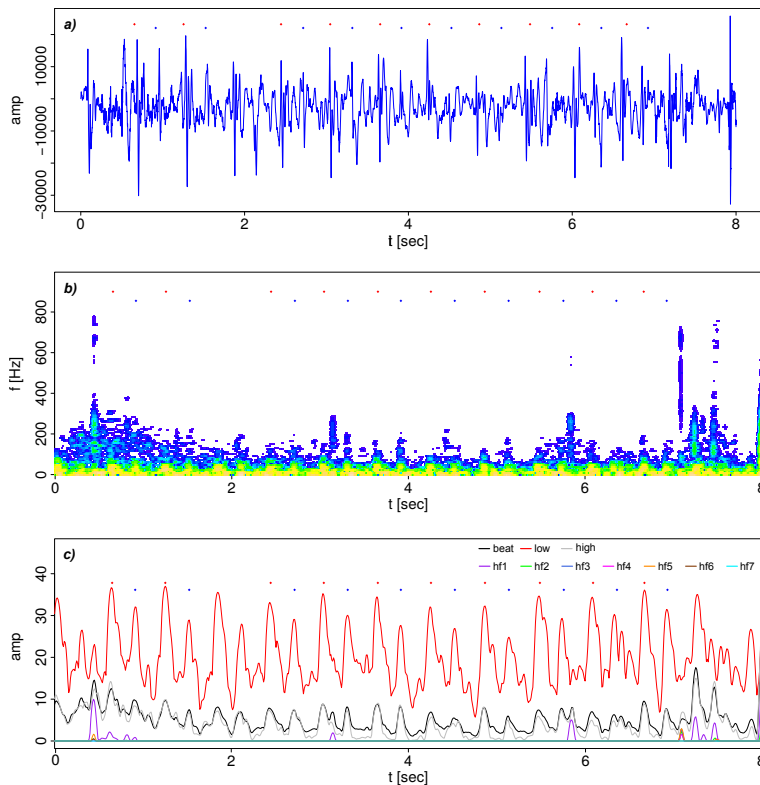
Figure 2. Example result for frequency analysis, beat and feature detection for PCG b0106 of the PhysioNet Challenge 2016 dataset: **a)** Raw waveform, **b)** spectrogram and **c)** frequency bands. The red and blue diamonds indicate S1 and S2, respectively.

5. Conclusion

We present tools for the analysis of phonocardiograms. Our official entry to the PhysioNet Challenge 2016 recognises abnormalities in PGCs with high sensitivity and accuracy. Still, automated diagnosis is not possible with the software, because too many abnormalities are detected in the recordings of healthy subjects as well.

Potential applicability of the software includes decision support for physicians: the software can identify potential abnormalities and display them as graphics or list. This way, the attention of the cardiologist is drawn to meaningful regions of the PCG. Responsibility for interpretation of the findings and for diagnosis stays with the physicians and must be based on their experience.

References

[1] Watrous RL. Computer-Aided Auscultation of the Heart: From Anatomy and Physiology to Diagnostic Decision Support. In 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS '06. 2006; 140–143.

[2] Naseri H, Homaeinezhad M, Pourkhajeh H. Noise/spike detection in phonocardiogram signal as a cyclic random process with non-stationary period interval. Comput Biol Med 2013;43:1205–1213.

[3] Kohonen T. The self-organizing map. Proc IEEE 1990; 78:1464–1480.

[4] Bergmeir C, Benítez JM. Neural networks in R using the stuttgart neural network simulator: RSNNS. J Stat Software 2012;46:1–26.

[5] Ari S, Kumar P, Saha G. A robust heart sound segmentation algorithm for commonly occurring heart valve diseases. J Med Ing Technol 2008;32:456–465.

[6] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiol Meas 2016;37.

[7] Sueur J, Aubin T, Simonis C. Seewave: a free modular tool for sound analysis and synthesis. Bioacoustics 2008;18:213–226.

[8] Krey S, Ligges U, Mersmann O. fftw: Fast FFT and DCT based on FFTW, 2011. R package version 1.0-3.

[9] Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth edition. New York: Springer, 2002. ISBN 0-387-95457-0.

Address for correspondence:

Andreas Dominik
THM, Wiesenstraße 14, D-35390 Gießen, Germany
andreas.dominik@mni.thm.de