Abnormal Heart Sounds Detected from Short Duration Unsegmented Phonocardiograms by Wavelet Entropy

Philip Langley[1], Alan Murray[2]

[1]University of Hull, Hull, UK
[2]Newcastle University, Newcastle, UK

Abstract

Segmentation of the characteristic heart sounds is thought to be an essential requirement for the automatic classification of phonocardiograms. The aim of this work was to test the feasibility of classification using short duration, unsegmented recordings.

Recordings from the 2016 PhysioNet/Computing in Cardiology Challenge were analysed. Wavelet entropy of unsegmented 5 s duration recordings was calculated and the optimum wavelet scale and wavelet entropy threshold determined from the training set. The algorithm was validated on the test set.

At a wavelet scale of 1.7 wavelet entropy was significantly reduced in abnormal recordings (median (IQR), 6.3 (1.8) vs 8.0 (1.8) p<0.0001). For a wavelet entropy threshold of 7.8 a score of 78% (sensitivity = 95%, specificity = 60%) was obtained on the training set. The robustness of this result was demonstrated on the test set which achieved a score of 77% (sensitivity = 98%, specificity = 56%).

Classification of unsegmented and short duration phonocardiograms by wavelet entropy is feasible.

1. Introduction

Heart sounds may indicate a range of cardiac pathologies such as vessel stenosis or valve regurgitation, with such abnormalities giving rise to different sound profiles to that of the normal, healthy heart [1].

The automated detection of abnormal heart sounds from electronic stethoscope recordings remains a significant challenge particularly when recordings are obtained outside the clinical environment or by novice users. Current classification algorithms typically segment and analyse specific heart sounds such as first (S1) and second (S2) heart sounds or the associated diastolic and systolic intervals [2]. However, segmentation of the heart sounds is in itself challenging, particularly in abnormal recordings where there is often less distinction between the individual heart sounds. Segmentation also increases the computational burden of such algorithms.

With this in mind the aim of this study was to assess the feasibility of using short, unsegmented recordings for phonocardiogram classification and to contribute to the 2016 PhysioNet/Computing in Cardiology Challenge [3]. Particularly it was thought that wavelet analysis might reveal differences in the time/frequency characteristics of normal and abnormal heart sounds.

2. Methods

This section first describes the methods of algorithm development followed by a description of the final algorithm. The dataset used in this study comprised a training set and test set and is described more completely elsewhere [2]. The algorithm was developed in the Matlab environment.

2.1. Algorithm development

Recordings were analysed by wavelet analysis to explore the time/frequency characteristics of the heart sounds. Using the 'Gaus4' mother wavelet the continuous wavelet transform coefficients were generated across a range of scales and translations according to

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^* \left(\frac{t-b}{a} \right) dt$$

where ψ^* is the complex conjugate of the wavelet function with scale and translation variables *a* and *b* respectively.

From the wavelet coefficients the wavelet energy at each scale and translation was calculated according to

$$E(a,b) = |T(a,b)|^2$$

Wavelet entropy, a measure of the temporal energy distribution as described in [4], was calculated according to

$$S(a) = - \int P(a,b) \log(P(a,b)) \, db$$

where the wavelet energy probability distribution was

defined as

$$P(a,b) = \frac{|T(a,b)|^2}{\int |T(a,b)|^2 db}$$

With this formulation wavelet entropy at scales with a broad temporal energy distribution is expected to be greater than those with a narrow temporal energy distribution.

Note that in the definition of wavelet entropy ($S(a)$) the integral is across the translation variable (b), which is equivalent to time in the time domain. Hence entropy will vary according to the length of the signal analysed regardless of its energy distribution. To avoid variations in entropy due to differences in recording lengths, the period analysed was fixed for all recordings. Figure 1 shows the histograms of recording durations for both the normal and abnormal recordings of the training set. As can be seen in figure 1 many of the recordings had lengths less than 8 s and the minimum was 5 s. By selecting a fixed analysis length of 5 s it was possible to analyse all the recordings of the training set without any exclusions. This was implemented by analysing only the first 5 s of a recording regardless of its overall length.
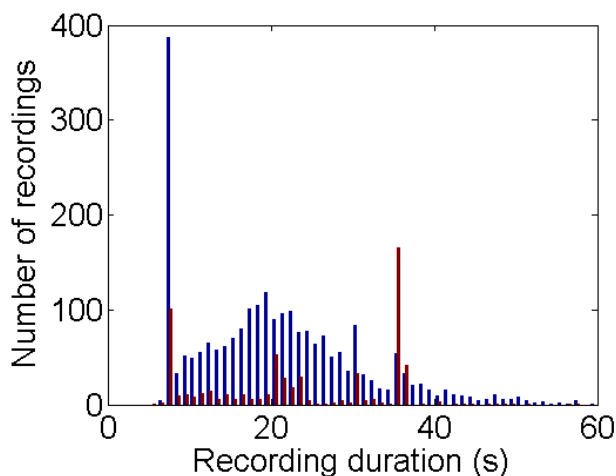

Figure 1. Histograms of the durations of the training set recordings classified as normal (light, blue) and abnormal (dark, red).

The optimal wavelet scale and wavelet entropy threshold for separating normal from abnormal recordings was the scale and threshold which achieved the highest score on the training set. The scoring algorithm was that originally proposed by the Challenge organisers defined as the average of specificity (Sp) plus sensitivity (Se)

$$score = (Se + Sp)/2$$

where

$$Se = \frac{Aa}{Aa + An}$$

$$Sp = \frac{Nn}{Na + Nn}$$

Aa, Nn represent correctly classified abnormal and normal recordings respectively and An and Na represent incorrectly classified abnormal and normal recordings respectively.

2.2. Final algorithm

The algorithm submitted to the official phase of the 2016 PhysioNet/Computing in Cardiology Challenge is illustrated in figure 2. Only the first 5 s of recording were analysed and wavelet analysis undertaken at the scale determined as optimal on the training set. Classification was based on the wavelet entropy threshold determined as optimum on the training set such that recordings with entropy exceeding the threshold were classified as 'Normal', otherwise they were classified as 'Abnormal'.
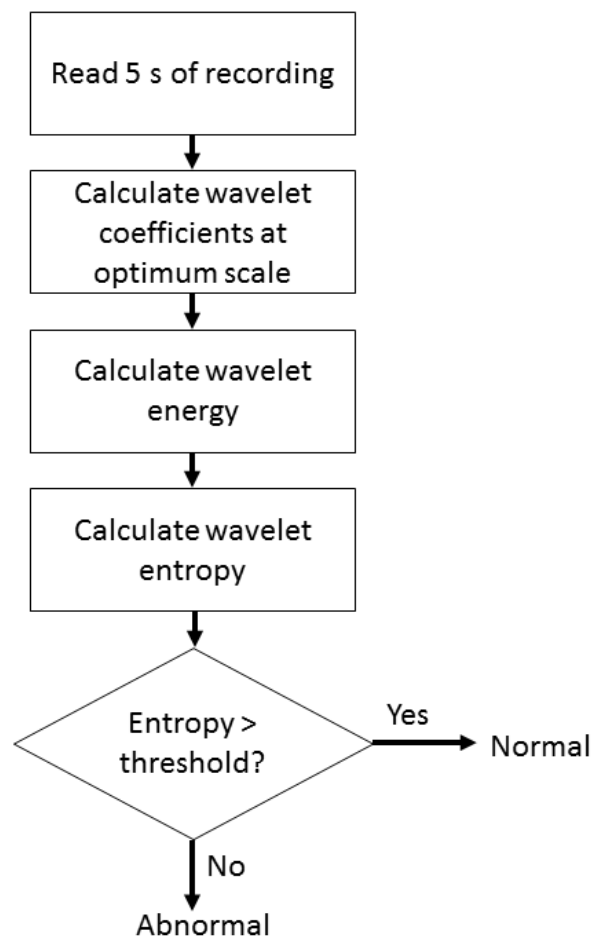

Figure 2. Flow chart to illustrate the processing steps of the classification algorithm.

3. Results

Figure 3 illustrates boxplots of wavelet entropy as a function of scale for all normal and abnormal recordings of the training set. The clearest distinction between groups was at scales in the range 1 to 3 (equivalent to a frequency range of 1000 to 333 Hz) and this distinction diminished with increasing scale.
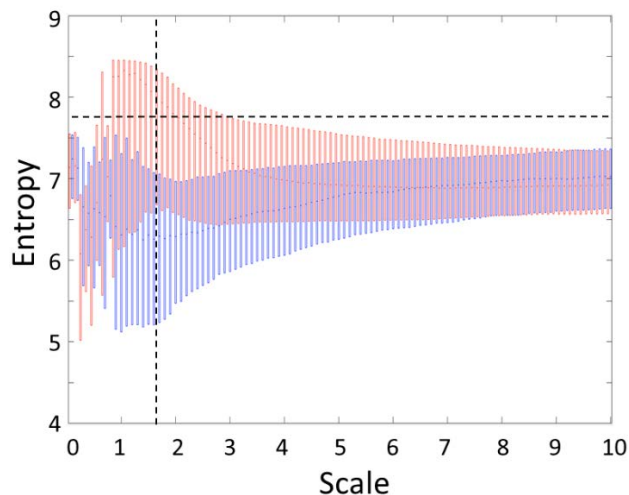

Figure 3. Boxplots of wavelet entropy as a function of wavelet scale for the training set. Normal recordings have red boxes and abnormal recordings have blue boxes. Boxplots show median and interquartile range. Optimal wavelet scale and entropy for separating normal and abnormal recordings are indicated by a vertical and horizontal dashed line respectively.

Optimal separation of normal and abnormal recordings was at wavelet scale 1.7. At this scale wavelet entropy was significantly less in the abnormal group than the normal group (median (IQR), 6.3 (1.8) vs 8.0 (1.8) $p<0.0001$). At this scale the optimal wavelet entropy threshold was 7.8 which achieved a training set score of 78% with high sensitivity (table 1). Reduced entropy at this scale suggests the presence of high frequency (~590 Hz) components in the heart sounds in abnormal recordings.

Table 1. Results of the classification algorithm on the training and test sets

Data set	Se (%)	Sp (%)	Score (%)
Training	95	60	78
Test	98	58	77

Figure 4 shows examples of correctly classified normal and abnormal recordings along with their wavelet coefficients, temporal energy distributions and entropy values. Similarly, examples of recordings incorrectly classified by the algorithm are shown in figure 5.
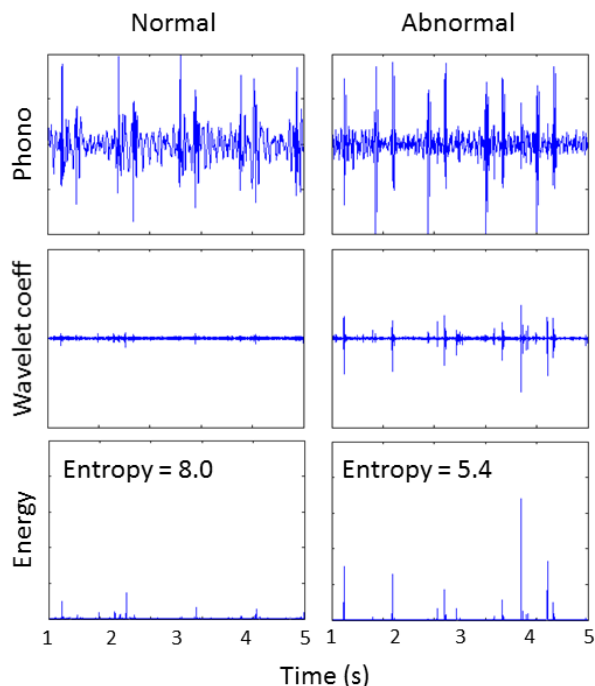

Figure 4. Examples of correctly classified normal (d0001) and abnormal (a0004) recordings. Phonocardiogram, wavelet coefficients and wavelet energy at scale 1.7 are shown. Amplitudes have been normalised across all plots.
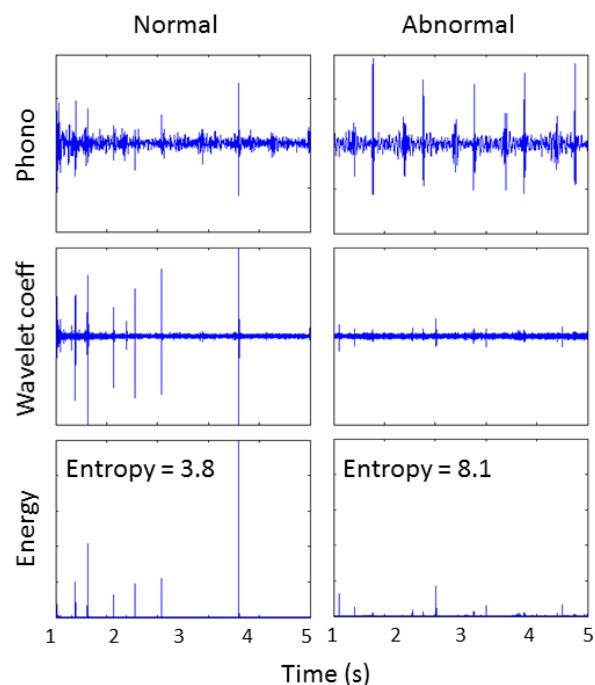

Figure 5. Examples of incorrectly classified normal (a0009) and abnormal (a0001) recordings. Phonocardiogram, wavelet coefficients and wavelet energy at scale 1.7 are shown. Amplitudes have been normalised across all plots.

There was no apparent fall in the performance of the algorithm when applied to the test set for which a score of 77% was achieved with high sensitivity. Table 1 compares the performance of the classification algorithm for training and test sets.

4.   Discussion

The feasibility of heart sound classification using short, unsegmented recordings has been demonstrated. Wavelet analysis revealed important differences in the time/frequency characteristics of the recordings and achieved promising classification performance with high sensitivity. The work holds promise for the development of computationally efficient algorithms which might be embedded into 'intelligent' devices to aid clinical decision making or for home use diagnostics.

Several important developments could be undertaken to improve the performance of the algorithm. Firstly, only a single feature of the phonocardiogram recordings was used. Second, no attempt was made to identify and exclude recordings with significant measurement noise. The addition of further features and exclusion of recordings with excessive noise would likely improve the classification accuracy.

References

[1] Nazeran H. Phonocardiography. In: Webster JG, editor. The Physiological Measurement Handbook. Boca Raton: CRC Press, 2015:101-21.

[2] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ et al. An open access database for the evaluation of heart sound algorithms. Physiol Meas 2016;37:in press.

[3] PhysioNet. Classification of normal/abnormal heart sound recordings: the PhysioNet/Computing in Cardiology Challenge 2016 [internet]. PhysioNet; 2016. Available from https://www.physionet.org/challenge/2016/

[4] Langley P. Wavelet entropy as a measure of ventricular beat suppression from the electrocardiogram in atrial fibrillation. Entropy 2015;17:6397-411.

Address for correspondence.

Philip Langley
School of Engineering and Computer Science
University of Hull
HU6 7RX
UK
p.langley@hull.ac.uk