Heart sound classification from unsegmented phonocardiograms

Title: Heart sound classification from unsegmented phonocardiograms

Philip Langley[1], Alan Murray[2]

[1] School of Engineering and Computer Science, University of Hull, Hull, UK
[2] School of Engineering, Newcastle University, Newcastle upon Tyne, UK

Abstract

Objective Most algorithms for automated analysis of phonocardiograms (PCG) require segmentation of the signal into the characteristic heart sounds. The aim was to assess the feasibility for accurate classification of heart sounds on short, unsegmented recordings.

Approach PCG segments of 5 second duration from the PhysioNet/Computing in Cardiology Challenge database were analysed. Initially the 5 second segment at the start of each recording (seg 1) was analysed. Segments were zero-mean but otherwise had no pre-processing or segmentation. Normalised spectral amplitude was determined by fast Fourier transform and wavelet entropy by wavelet analysis. For each of these a simple single feature threshold based classifier was implemented and the frequency/scale and thresholds for optimum classification accuracy determined. The analysis was then repeated using relatively noise free 5 s segments (seg 2) of each recording. Spectral amplitude and wavelet entropy features were then combined in a classification tree.

Main results There were significant differences between normal and abnormal recordings for both wavelet entropy and spectral amplitude across scales and frequency. In the wavelet domain the differences between groups were greatest at highest frequencies (wavelet scale 1, pseudo frequency 1 kHz) whereas in the frequency domain the differences were greatest at low frequencies (12 Hz). Abnormal recordings had significantly reduced high frequency wavelet entropy: (Median (interquartile range)) 6.63 (2.42) vs 8.36 (1.91), $p < 0.0001$, suggesting the presence of discrete high frequency components in these recordings. Abnormal recordings exhibited significantly greater low frequency (12 Hz) spectral amplitude: 0.24 (0.22) vs 0.09 (0.15), $p< 0.0001$. Classification accuracy (mean of specificity and sensitivity) was greatest for wavelet entropy: 76% (specificity 54%, sensitivity 98%) vs 70% (specificity 65%, sensitivity 75%) and was further improved by selecting the lowest noise segment (seg 2): 80% (specificity 65%, sensitivity 94%) vs 71% (specificity 63%, sensitivity 79%). Classification tree with combined features gave accuracy 79% (specificity 80%, sensitivity 77%).

Significance The feasibility of accurate classification without segmentation of the characteristic heart sounds has been demonstrated. Classification accuracy is comparable to other algorithms but achieved without the complexity of segmentation.

1. Introduction

Heart sounds have long been recognised as a diagnostic tool which can indicate a range of cardiac pathologies related to valve disease (Nazeran 2015). Heart sounds are generated by turbulent flow of blood and the resultant mechanical vibrations are transmitted through the torso and can be heard at different locations on the body surface using a stethoscope. Recordings of the heart sounds in electronic format via an electronic stethoscope enable the processing and analysis of the sounds with the potential for automated diagnosis (Brusco & Nazeran 2006). Many works have looked at this challenging problem using different signal processing methods. For example, Maglogiannis et al (2009) used wavelet decomposition for heart sound segmentation and support vector machine for classification and reported accuracy of 91% for classification of 'normal' and 'pathological' heart sounds. Where available simultaneously recorded ECG can improve diagnostic performance by providing definitive cardiac cycle reference points. For example, El-Segaier et al (2005) use short time Fourier transform and associated frequency characteristics of the heart sounds for both segmentation and classification aided by ECG reference points. A review of segmentation and classification methods is provided in Liu et al (2016). Classification of heart sounds using only PCGs was recently posed as the PhysioNet/Computing in Cardiology Challenge 2016 (PhysioNet 2016, Goldberger 2000) and attracted a large number of entries (Clifford *et al* 2016). Almost universally the proposed algorithms performed segmentation of the recording into the characteristic heart sounds S1, S2 and associated systolic and diastolic intervals. While such segmentation provides many classification features which may be useful in identifying abnormal heart sounds it also introduces considerable complexity and increased computational burden into the algorithms (Schmidt *et al* 2010). Here the aim was to test the feasibility of accurate classification of heart sounds from recordings without segmentation of heart sounds and intervals. Following our initial work based on classification of heart sounds using wavelet entropy of unsegmented recordings (Langley & Murray 2016) here we extend the work to include spectral amplitude as a classification feature and the selection of noise free segments.

2. Methods

2.1 Database of recordings

Data for this study was obtained from the PhysioNet/Computing in Cardiology Challenge 2016 (PhysioNet 2016). The dataset is fully described in Liu et al (2016) and the Challenge is fully described in Clifford et al (2016). It comprises a training set of PCG recordings of variable duration classified as either 'normal' (2408 recordings) or 'abnormal' (630 recordings) and a hidden test set which was unavailable to challenge participants. Recordings classified as normal were from healthy subjects and those classified as abnormal from patients with a confirmed cardiac diagnosis. The database is heterogeneous since it contains recordings from several contributing centres, with no standardised recording position, from adults and children and using different recording instruments (Liu *et al* 2016). Sample rate was 2000 Hz for all recordings. Recordings had an additional descriptor 'clean' or 'noisy,' but this was not utilised in the present study since we employed our own assessment of recording noise. Since all recordings in the training set had either the classification of 'normal' or 'abnormal' the aim of our study was to make the binary classification 'normal' or 'abnormal' without use of a third 'uncertain' classification category. All analysis was done in the Matlab environment using the appropriate tool boxes.

2.2. Recording duration

The database comprises recordings of variable duration. Although the recording duration of some phonocardiograms extended to almost 120 s the vast majority had short duration of less than 8 s with the shortest duration of 5 s (Langley & Murray 2016, Liu *et al* 2016). Hence to fix on a consistent analysis length for all recordings in this study the analysed data length was 5 s for all recordings. Initially the segment analysed was the first 5 s segment of each recording (seg 1). However, some recordings had considerable noise at the start of the recordings so the analysis was repeated on 5 s segments with lowest noise (seg 2). The selection of this segment is described in the appropriate section 2.4.

2.3 Classification features

Training set recordings were analysed by both spectral and wavelet techniques to explore the time/frequency characteristics of the heart sounds. Specifically, two features i) spectral amplitude and ii) wavelet entropy were assessed as features for classification of heart sounds. Spectral amplitude measures the relative amplitude of the signal as a function of frequency. Wavelet entropy on the other hand measures the temporal energy distribution as a function of frequency. The presence of abnormal heart sounds, such as murmurs, clicks and rubs was expected to generate distinct spectral amplitude and wavelet entropy characteristics compared to normal heart sounds. In the following sections the calculation of these characteristics is described along with their use as classification features.

2.3.1 Spectral amplitude

After subtracting the mean from the 5 s segment the one-sided amplitude spectrum was calculated using fast Fourier transform (Rao *et al* 2011). The amplitude was normalised to the peak amplitude. The segment length (10000 samples (5 s)) provided a spectral resolution of 0.2 Hz across the range 0 to 1000 Hz. This provided spectral amplitudes in 5001 frequency bins (0 to 1000 Hz) for each recording. According to the observed differences between groups in spectral amplitude across the full frequency range, a simple threshold based classification algorithm was implemented to assign recordings to either ‘normal’ or ‘abnormal’ groups depending the recording’s spectral amplitude relative to a threshold. Specifically, the spectral amplitude of abnormal recordings was found to be significantly greater than normal recordings at low frequencies (see results section for full details) so the classification algorithm assigned recordings to the ‘abnormal’ class if the spectral amplitude at a given frequency was greater than the threshold and to the ‘normal’ class otherwise. This algorithm was implemented for all frequency bins and the threshold yielding the highest classification accuracy at each bin was determined by sequentially incrementing the threshold from a base value and identifying the threshold achieving the greatest accuracy. Finally, with the aim of producing a single feature classifier, the frequency bin (and associated threshold) yielding the highest classification accuracy was selected as the optimum. A colour map of classification accuracy was plotted as a function of frequency and threshold.

2.3.2 Wavelet entropy

Using the 'Gaus4' mother wavelet the continuous wavelet transform coefficients were generated according to

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

where $\psi^*$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

$$E(a,b) = |T(a,b)|^2 \tag{2}$$

Wavelet entropy, a measure of the temporal energy distribution, was calculated according to (Langley 2015)

$$S(a) = -\int P(a,b) \log\big(P(a,b)\big)\, db \tag{3}$$

where the wavelet energy probability distribution was defined as

$$P(a,b) = \frac{|T(a,b)|^2}{\int |T(a,b)|^2 db} \tag{4}$$

With this formulation wavelet entropy is calculated at each scale. Scales with a broad temporal wavelet energy distribution, (ie a scale with energy distributed across the duration of the recording segment) would have greater wavelet entropy than scales with a narrow temporal energy distribution (ie a scale with energy concentrated at particular time points of the recording segment) (Langley 2015). It was expected that the presence of heart sound abnormalities such as clicks and rubs would have temporally concentrated energy at distinct scales resulting in reduced wavelet entropy at those scales. Note that wavelet entropy is calculated using the wavelet energy probability distribution (equation 4) and as such it is not dependent upon the amplitude of the wavelet coefficient. Hence it characterises the temporal energy distribution even at scales with low amplitude.

Wavelet entropy was calculated for scales 1 to 100 with increment 0.1 which, for the 'Gaus4' mother wavelet, corresponded to the frequency range 1000 Hz to 10 Hz. Scales below 1 were not considered as they correspond to frequencies above the Nyquist frequency. Similar to the analysis of the spectral amplitude, a simple threshold based classification algorithm was implemented to assign recordings to either 'normal' or 'abnormal' groups depending on the recording's wavelet entropy relative to a threshold. Particularly it was noted that wavelet entropy was greatly reduced for abnormal recordings at the lowest scales (highest frequencies) so the classification algorithm assigned recordings to the 'normal' class if the wavelet entropy at a given scale was greater than the threshold and to the 'abnormal' class otherwise. To determine the scale and threshold yielding the highest classification accuracy a colour map of classification accuracy was plotted as a function of scale and threshold.

2.3.3 Combined spectral amplitude and wavelet entropy classifier

Having developed two single feature classifiers based on either spectral amplitude or wavelet entropy alone as described in sections 2.3.1 and 2.3.2, the utility of using both these features in a single classifier was assessed. A classifier taking both wavelet entropy and spectral amplitude as classification features was designed using a decision tree. The decision tree approach was used because overtraining of the classifier can be avoided by limiting the number of decision nodes. Also, the resulting classifier retains physical meaning of classification features so is simple to interpret. As such we designed a decision tree classifier having the minimum number of decision nodes using the Matlab 'fitctree' command trained on the training set data.

2.4 Selection of lowest noise segment of recordings

It was noticed that some recordings had significant noise during the first 5 s segment of the recording while other parts of the recordings were relatively noise free. To identify the segments with lowest noise levels to use in the analysis as an alternative to the first 5 s segment it was necessary to estimate the noise content of segments. PCG signal noise due to patient movement or other recording artefact generally produces large amplitude disturbance in the signal. An example is shown in figure 1. Wavelet entropy described in the previous section was used as a relative measure of signal noise within recordings. Noisy recording segments have a concentration of signal energy during the noise disturbance which is reflected in low wavelet entropy. For clean recording segments signal energy occurs more evenly temporally distributed across the recording (with energy peaks only at each heart sound) and this is reflected in higher wavelet entropy. In another application wavelet entropy was used previously as a measure of noise due to residual ventricular activity in electrocardiogram recordings of the abnormal heart rhythm atrial fibrillation (Langley 2015). As an illustrative example consider figure 1. Figure 1 top panel shows a PCG recording exhibiting considerable noise during the first few seconds while the later parts are relatively noise free. Figure 1 lower panels show the wavelet entropy for the noisy and clean segments. It indicates that noise free segments had higher wavelet entropy across all scales. In order to automatically select the cleanest segment of a recording, wavelet entropy was calculated for 5 s segments in increments of 1 s intervals across the full recording and the segment providing the highest entropy was selected as the lowest noise segment (seg 2). Spectral and wavelet analysis was repeated on these 5 s segments.

2.5 Statistical analysis

Specificity and sensitivity of classification were calculated according to standard formula by counting the number of true/false positive/negative classifications on the training set. A measure of classification accuracy was defined as (specificity + sensitivity)/2 as in the PhysioNet/Computing in Cardiology Challenge 2016 (Clifford *et al* 2016). Cross-validation was implemented by bootstrapping 300 random samples of the training set recordings over 1000 iterations and mean and standard deviation of specificity, sensitivity and accuracy are reported. Data were non-normally distributed so significance of differences between features of the normal and abnormal recordings of the training set were evaluated with the Wilcoxon rank sum test.

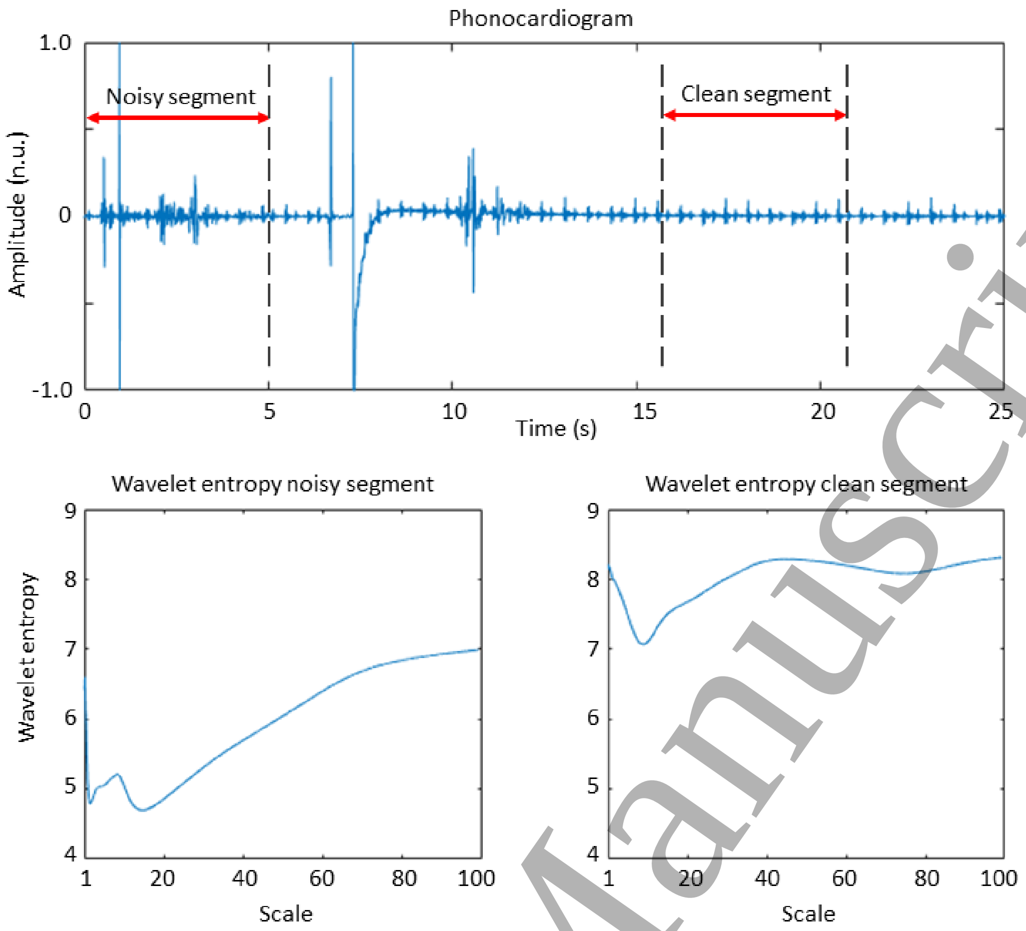

Figure 1. Illustration of a PCG recording with noisy and clean segments and their associated wavelet entropies. The first 5 s of the recording is contaminated by noise resulting in low wavelet entropy (mean wavelet entropy = 6). The cleanest 5 s segment had higher wavelet entropy (mean wavelet entropy = 8).

3. Results

For the single feature classifiers the results obtained by analysis of the first 5 s segment of each recording are presented in sections 3.1 and 3.2. Section 3.3 presents the results for the lowest noise segments and the decision tree classifier.

3.1 Spectral amplitude

The median normalised amplitude spectra for normal and abnormal recordings are shown in figure 2 (left panel). There were clear differences in the median spectral amplitude with abnormal recordings exhibiting higher amplitude over the range 0 to 35 Hz. Based on these distributions of amplitude a simple classifier which assigned recordings to the 'abnormal' group for spectral amplitude above 'threshold' and to the 'normal' group for those equal or below 'threshold' was implemented. The classification accuracy map for this algorithm showing the accuracy of classification as a function of frequency and threshold is shown in the right panel of figure 2. The highest accuracy was 70% (specificity 65%, sensitivity 75%) at a frequency of 12 Hz with a threshold of 0.14.

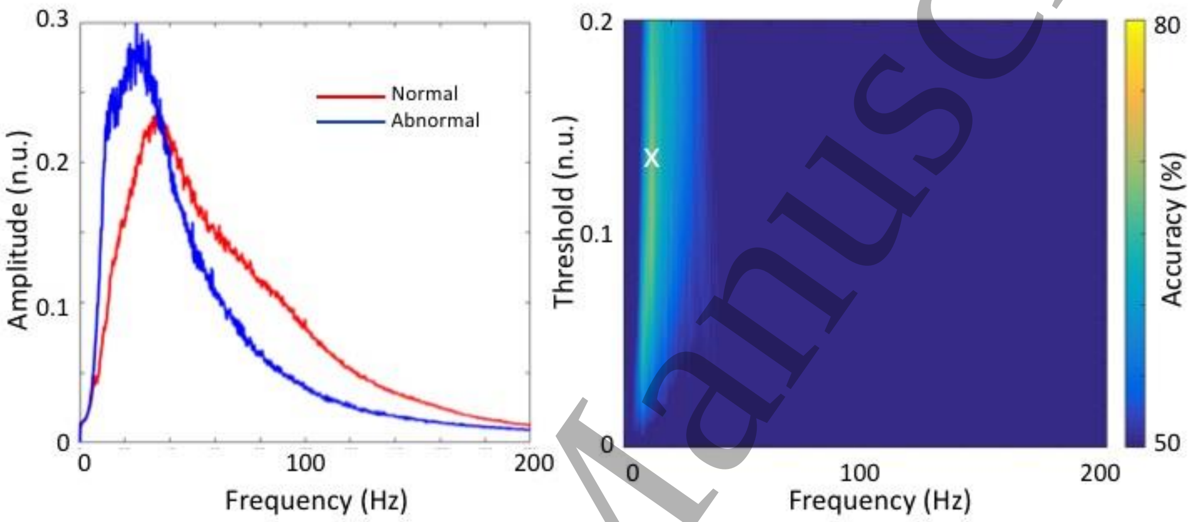

Figure 2. Left panel: Median spectral amplitudes from fast Fourier transform analysis of normal (red) and abnormal (blue) PCG recordings. Right panel: Classification accuracy map for a simple classifier (spectral amplitude of abnormal recordings greater > threshold) showing greatest accuracy (70%) at 12 Hz for a threshold of 0.14 (white cross).

Statistical analysis of the differences between spectral amplitude for normal and abnormal recordings at 12 Hz showed that median (interquartile range) amplitude was significantly greater in the abnormal group compared to the normal group (0.24 (0.22) vs 0.09 (0.15), $p < 0.0001$) as illustrated in figure 3.

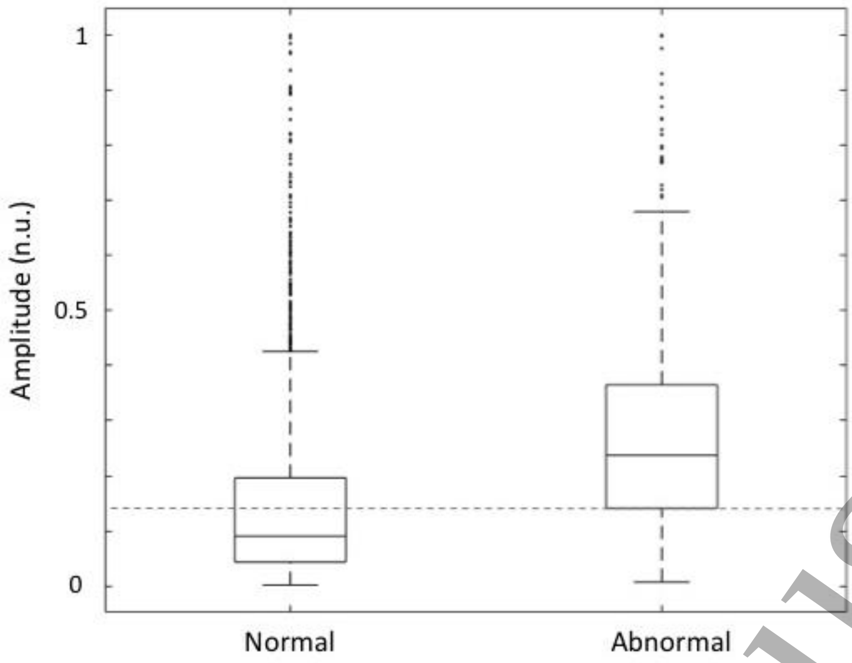

Figure 3. Spectral amplitude distributions at 12 Hz for normal and abnormal recordings. Abnormal recordings had significantly greater amplitude at this frequency. The dashed line indicates the threshold (0.14) for maximum classification accuracy at this frequency.

Figure 2 left panel also indicates that median amplitude was lower for abnormal recordings above 35 Hz, however, a simple classifier based on assigning recordings to the abnormal group if their amplitude was below 'threshold' yielded a maximum accuracy of 66% (at 100 Hz) so was not considered further.

3.2 Wavelet entropy

Figure 4 left panel shows median wavelet entropy across scales from 1 to 30 for both the normal and abnormal heart sounds. Although median entropy was greater for abnormal recordings at scales greater than 7 the largest difference between median entropy was at the lowest wavelet scales where the entropy of normal heart sounds exceeded those of abnormal heart sounds. Based on these distributions of wavelet entropy a simple classifier which assigned recordings to the 'abnormal' group for wavelet entropy below 'threshold' and to the 'normal' group for those equal or above 'threshold' was implemented. The classification accuracy map for this algorithm showing the accuracy of classification as a function of scale and threshold is shown in the right panel of figure 4. Accuracy was highest at scales from 1 to 2 with corresponding pseudo frequencies from 1000 to 500 Hz respectively. The highest accuracy 76% (specificity 54%, sensitivity 98%), was obtained for scale 1 with a threshold of 8.3.

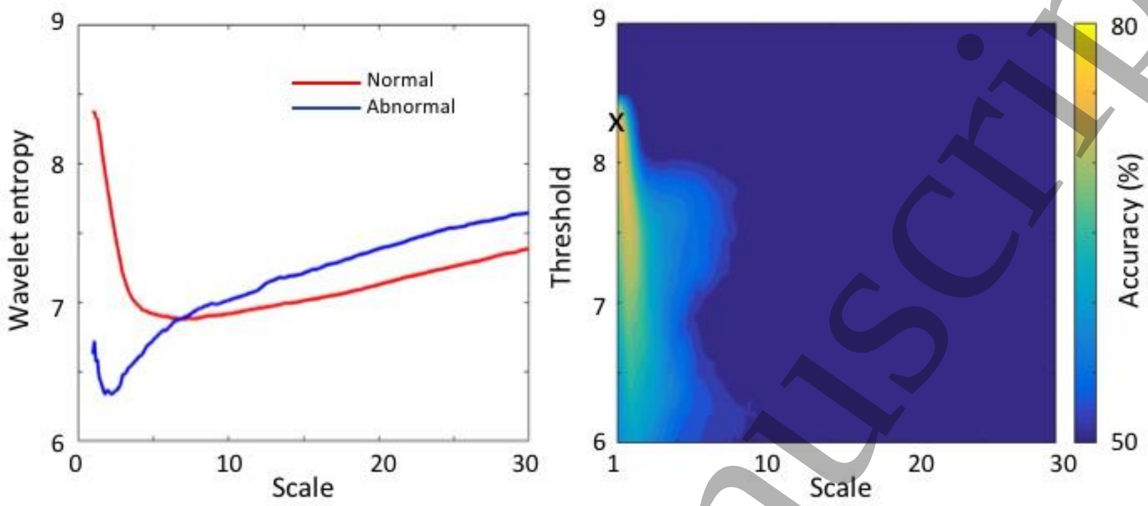

Figure 4. Left panel: Median wavelet entropy from wavelet analysis of normal (red) and abnormal (blue) PCG recordings. Right panel: Classification accuracy map for a simple classifier (wavelet entropy of abnormal recordings less than threshold) showing greatest accuracy (76%) at scale 1 for a threshold of 8.3 (black cross).

Median (interquartile range) wavelet entropy was significantly greater in the normal heart sound recordings at this scale (8.4 (1.9) vs 6.6 (2.4) $p < 0.0001$) as illustrated in figure 5. Wavelet entropy for the normal group were highly negatively skewed (figure 5) which is thought to be due to the poor quality of some of the recordings.

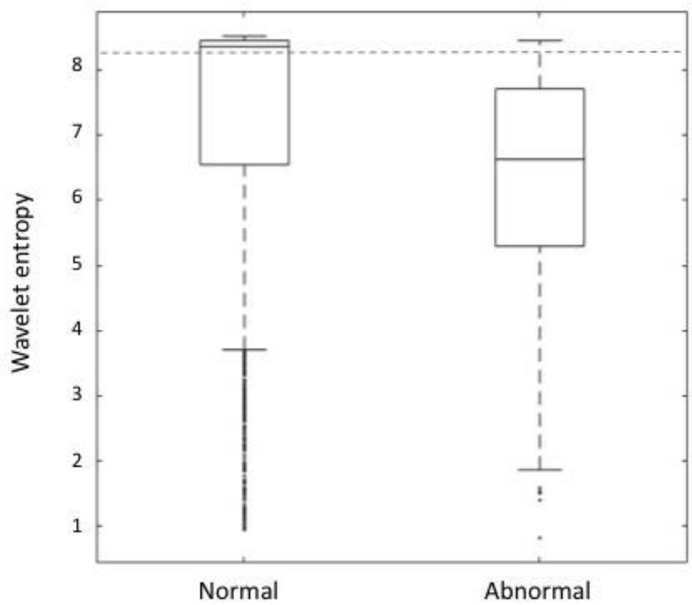

Figure 5. Wavelet entropy distributions at scale 1 for normal and abnormal recordings. Normal recordings had significantly greater wavelet entropy at this scale. The dashed line indicates the threshold (8.3) for maximum classification accuracy at this scale.

3.3 Spectral amplitude and wavelet entropy for lowest noise segments

The data presented so far relate to the 5 s segment at the start of the recording (seg 1). The analysis was repeated for the 5 s segments exhibiting the lowest noise (seg 2) and the data are presented here.

As expected wavelet entropy increased for both normal (seg 1: 8.36 (1.91) vs seg 2: 8.42 (0.69)) and abnormal (seg 1: 6.63 (2.42) vs 7.50 (1.63)) recordings as a result of the automatic selection of the lowest noise segments (table 1). Classification accuracy increased from 76% (seg 1) to 80% (seg 2) when applying the wavelet entropy algorithm to the lowest noise segments with no change in optimum scale or threshold (table 1). Spectral amplitude was relatively unaffected by the choice of segments and accuracy only increased marginally. Table 1 compares the performance of classification by spectral amplitude and wavelet entropy for both segments.

Table 1. Spectral amplitude and wavelet entropy parameters and classification performance for the first 5 s segments (seg 1) and lowest noise segments (seg 2) on the full training set.

Spectral Amplitude							
	Freq (Hz)	Thres	Se (%)	Sp (%)	Acc (%)	Abnormal Median (IQR)	Normal Median (IQR)
seg 1	12.0	0.14	75	65	70	0.24 (0.22)	0.09 (0.15)
seg 2	11.4	0.12	78	63	71	0.24 (0.24)	0.09 (0.14)
Wavelet entropy							
	Scale	Thres	Se (%)	Sp (%)	Acc (%)	Abnormal Median (IQR)	Normal Median (IQR)
seg 1	1	8.3	98	54	76	6.63 (2.42)	8.36 (1.91)
seg 2	1	8.3	94	65	80	7.50 (1.63)	8.42 (0.69)

Figure 6 provides illustrative examples of normal and abnormal PCG recordings and their associated spectral amplitude, wavelet energy and wavelet entropy distributions. Their wavelet coefficients at scale 1 are also shown. These examples were chosen because their spectral amplitudes at 12 Hz (0.09 (normal) vs 0.25 (abnormal)) and wavelet entropy at scale 1 (8.4 (normal) vs 6.7 (abnormal)) were close to the group median values so are representative examples of the normal and abnormal recordings. Note that the abnormal recording exhibits spikes in its wavelet coefficients at scale 1 (figure 6 (row E)) resulting in low wavelet entropy at this scale. The normal recording does not exhibit such spikes resulting in a larger wavelet entropy at this scale.

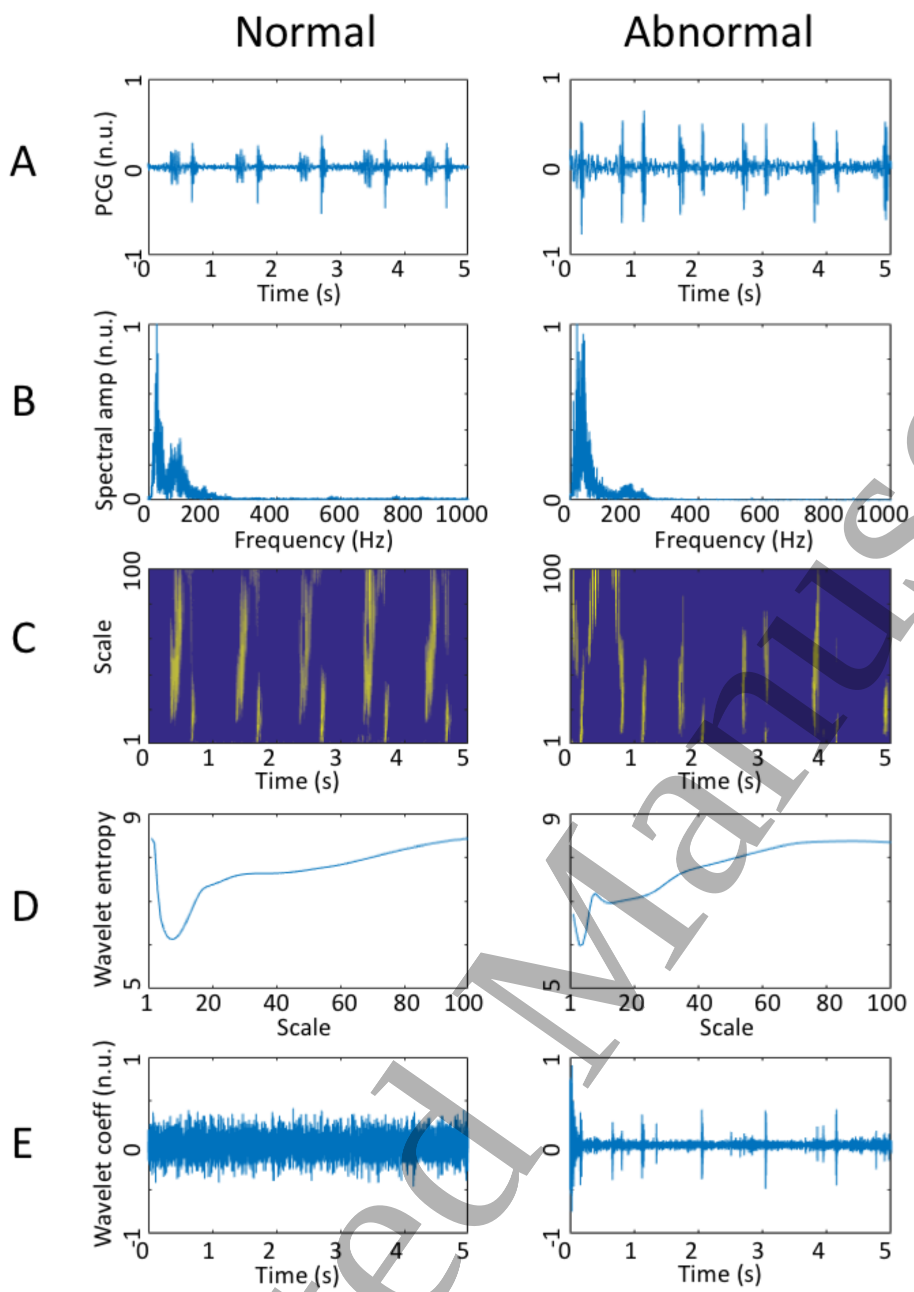

Figure 6. Example normal and abnormal PCG recordings (row A) and their spectral amplitude distribution (row B), wavelet coefficient energy distribution (row C) (light colour indicates maximum energy), wavelet entropy distribution (row D) and wavelet coefficients for wavelet scale 1 (row E).

3.4 Combined feature classifier

The decision tree used to combine the spectral amplitude (12 Hz) and wavelet entropy (scale 1) features is illustrated in figure 7. The minimum number of decision nodes was 3. All recordings with wavelet entropy of 8.3 or greater were classified as normal consistent with the threshold determined for the single feature classifier. Only recordings with wavelet entropy between 5.6 and 8.3 and spectral amplitude of 0.07 or greater were classified as abnormal. Although classification accuracy was no better than the single feature wavelet entropy classifier at 79%, this was achieved at higher specificity (80%) than either of the single feature classifiers. Sensitivity was 78%.

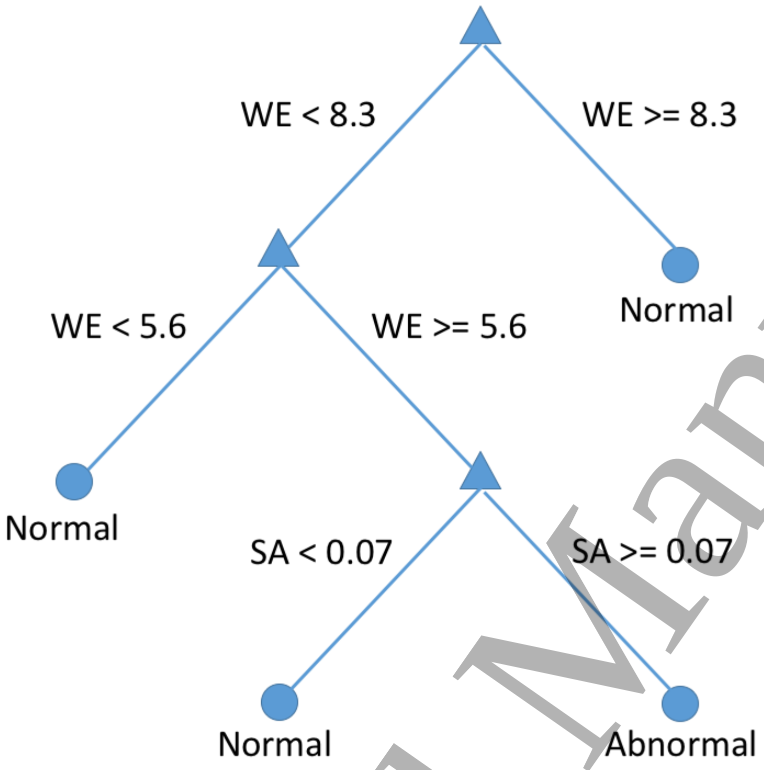

Figure 7. Decision tree for the two feature classifier with the minimum number of decision nodes (n = 3). The two features are wavelet entropy at scale 1 (WE) and spectral amplitude at 12 Hz (SA).

3.5 Cross validation

Results from the cross validation study for each of the single feature classifiers and the combined feature decision tree classifier are presented in table 2. Cross validation was only performed on the lowest noise segments (seg 2). Mean values are very similar to those obtained from the entire training set demonstrating the robustness of the algorithms. Standard deviations of performance measures were similar for all classifiers with the exception of the sensitivity of the single feature wavelet entropy classifier which had half the variability of the other classifiers.

Table 2. Classification performance from cross-validation. Values are mean (standard deviation) across 1000 bootstrap iterations.

Classifier	Se (%)	Sp (%)	Acc (%)
Spectral amplitude	75 (6)	60 (3)	68 (3)
Wavelet entropy	94 (3)	65 (3)	80 (2)
Decision tree	77 (5)	80 (3)	79 (3)

Further validation was provided by our initial wavelet entropy algorithm submitted as an entry to the PhysioNet Challenge which achieved a score of 76% (specificity 56%, sensitivity 96%) on the Challenge test set (Langley & Murray 2016). This demonstrates consistent performance across both training and test sets of the PCG database. We were unable to submit further entries to evaluate the performance of subsequent algorithms on the test set.

4. Discussion

Two single feature algorithms for classification of short unsegmented PCGs have been demonstrated. Using features of either low frequency spectral amplitude or low scale wavelet entropy, a simple threshold classifier achieved accuracies of greater than 70%. Of the two algorithms wavelet entropy proved to be the best performing with up to 10% improved accuracy with high sensitivity (> 94%) compared to spectral amplitude (table 1). Combining these classification features into a decision tree classifier resulted in similar classification accuracy but with reduced sensitivity (78%) and increased specificity (80%).

Abnormal recordings had significantly higher spectral amplitude at low frequencies compared to normal recordings, consistent with the presence of low frequency murmurs in abnormal recordings. The differences in spectral amplitude were most significant at frequencies around 12 Hz and this frequency yielded classification accuracies of around 70%. Note that this frequency is below the human audible frequency range so would be unlikely to be detected by manual auscultation. Wavelet entropy proved to be the better performing algorithm. It was shown that by selecting the 5 s segment with the highest wavelet entropy as a measure of the lowest noise segment the classification accuracy was improved from 76% to 80% on the training set. Abnormal recordings were associated with reduced wavelet entropy at the lowest scales. This suggests the presence of discrete high frequency (500 to 1000 Hz) components in abnormal recordings. Low amplitude, high frequency components such as this have been noted previously, in particularly with murmurs associated with regurgitation (Leatham 1975, Liu *et al* 2016).

The unique feature of the approach is that feature extraction is performed without the requirement for segmentation of the recordings into the characteristic heart sounds and systolic and diastolic intervals. This has the potential to significantly reduce the complexity and computational burden of the algorithms and facilitate their implementation as embedded algorithms in PCG devices. For example, we tested the execution time to classify a recording by the unsegmented single feature wavelet entropy classifier compared to the sample logistic regression-hsmm heart sound segmentation based classifier available on PhysioNet (PhysioNet 2016). The unsegmented classifier executed on average 11 times faster than the segmented one. It might have been expected that this simple approach would yield considerably poorer classification performance compared to algorithms using segmented recordings. However, the wavelet entropy algorithm showed comparable accuracy to other algorithms in the Computing in Cardiology/PhysioNet Challenge and was ranked 34th out of 47 entries and achieved a score of 76% which was slightly lower than the median (range) score of 79% (54 – 86%) of all the entries submitted to the test set.

It should be noted that the PCG database is comprised of recordings from multiple centres (Liu *et al* 2016). It was noted that there were considerable differences between spectral amplitude and wavelet entropy characteristics between the recordings from different centres. So although our algorithms are based on features derived from the recordings from all the centres, caution must be used when applying the algorithms to new data. It is however reassuring that the wavelet entropy algorithm performed comparably on training and test sets, especially since the test set contained recordings from two centres not included in the training set (Liu *et al* 2016).

In conclusion the feasibility of accurate classification without segmentation of the characteristic heart sounds has been demonstrated. Classification performance is comparable to other algorithms but achieved without the complexity of segmentation.

References

Brusco M and Nazeran H 2006 Development of an intelligent PDA-based wearable digital phonocardiograph. *In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (pp. 3506-3509). IEEE.

Clifford G D, Liu C Y, Moody B, Springer D, Silva I, Li Q and Mark R G 2016 Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016. *Comput. Cardiol*. **43** DOI:10.22489/CinC.2016.179-154.

El-Segaier M, Lilja O, Lukkarinen S, Sörnmo L, Sepponen R and Pesonen E 2005 Computer-based detection and analysis of heart sound and murmur. *Ann. Biomed. Eng*. **33** 937-942

Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 Physiobank, physiotoolkit, and physionet. *Circulation* **101** e215-e220

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

Langley P, Murray A 2016 Abnormal Heart Sounds Detected from Short Duration Unsegmented Phonocardiograms by Wavelet Entropy, *Comput. Cardiol.* 2016 **43** DOI:10.22489/CinC.2016.156-268

Leatham A 1975 Auscultation of the Heart and Phonocardiography (London: Churchill Livingstone)

Liu C, Springer D, Li Q, Moody B, Juan R A, Chorro F J, Castells F, Roig J M, Silva I, Johnson A E and Syed Z 2016 An open access database for the evaluation of heart sound algorithms. *Physiol. Meas*. **37** 2181-2213

Maglogiannis I, Loukis E, Zafiropoulos E and Stasis A 2009 Support vectors machine-based identification of heart valve diseases using heart sounds. *Comput. Methods Programs Biomed*. **95** 47-61

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

PhysioNet 2016 Classification of normal/abnormal heart sound recordings Available online: https://www.physionet.org/challenge/2016 [accessed 20/1/2017]

Rao K R, Kim D N and Hwang J J 2011 Fast Fourier transform-algorithms and applications. Springer Science & Business Media

Schmidt S E, Holst-Hansen C, Graff C, Toft E and Struijk J J 2010 Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiol. Meas*. **31** 513-29