A Robust Heart Sound Segmentation and Classification Algorithm using Wavelet Decomposition and Spectrogram

Yiqi Deng
Dept. of Computer Science, UCL
Malet Place,
London WC1E 6BT
y.deng.11@ucl.ac.uk

Peter J Bentley
Dept. of Computer Science, UCL
Malet Place,
London WC1E 6BT
p.bentley@cs.ucl.ac.uk

Abstract

This short article summarizes UCL's entry for the PASCAL Classifying Heart Sounds Challenge. The approach focused on the creation of novel segmentation and classification methods based on wavelet decomposition and spectrogram analysis.

1 Introduction

The Classifying Heart Sounds Challenge aims to achieve preliminary screening of cardiac pathologies by analyzing the features of heartbeat collected from digital stethoscope and mobile devices.

According to the World Health Organization, cardiovascular diseases (CVDs) are the number one cause of death globally: more people die annually from CVDs than from any other cause. An estimated 17.1 million people died from CVDs in 2004, representing 29% of all global deaths. Of these deaths, an estimated 7.2 million were due to coronary heart disease [1]. Any method which can help to detect signs of heart disease could therefore have a significant impact on world health.

The challenge is to produce methods to do exactly that. Specifically, we are interested in creating the first level of screening of cardiac pathologies both in a Hospital environment by a doctor (using a digital stethoscope) and at home by the patient (using a mobile device).

The problem is of particular interest to machine learning researchers as it involves classification of audio sample data, where distinguishing between classes of interest is non-trivial. Data is gathered in real-world situations and frequently contains background noise of every conceivable type. The differences between heart sounds corresponding to different heart symptoms can also be extremely subtle and challenging to separate. Success in classifying this form of data requires extremely robust classifiers. Despite its medical significance, to date this is a relatively unexplored application for machine learning.

The challenge consists of two tasks: Heart Sound Segmentation and Heart Sound classification. The first task is to produce a method that can locate the first heart sound S1(lub) and the second heart sound S2(dub) and segment Normal audio files from two datasets A and B (described later). The second task aims at classifying the heartbeat audio into one of the four categories for Dataset A (Normal, Murmur, Extra Heart Sound and Artifact) and three categories for Dataset B (Normal, Murmur and Extrasystole).

2 Background

Some attempts to segment phonocardiographic (PCG) signals have been reported in the literature. The majority of them exploit electrocardiogram (ECG) signals or/and carotid pulse data. For example, Groch presented a solution where the segmentation was based on the time domain characteristics of the signal [2]. Strunic extracted signals on certain band to reduce anomalies and then set a amplitude threshold to pick out the spikes and realize the segmentation [3]. To achieve classification, Karraz extracted the QRS complex from the signal as features and applied them into a Neural Network Classifier based on a Bayesian framework. Spencer integrated all the segmented heart cycles into one average heart cycle and used it to train the Artificial Neural Network (ANN) to classify heartbeat into Normal, Systolic Murmur caused by Mitral Regurgitation (MR), Systolic Murmur caused by Aortic Stenosis (AS) and Diastole Murmur caused by Aortic Regurgitation (AR) [3]. According to his result, when processing simulated heart sounds, the accuracy and sensitivity of ANN could be as high as $76\pm6.1\%$ and $89.7\pm5.9\%$ respectively. The accuracy drops to $48.7\pm12.7\%$ when he used data collected by an electronic stethoscope with a duration of about 5 seconds. In attempts to apply this system to the challenge datasets, for most cases it reported error because it could not deal with audio files of largely different lengths. When provided with selected audio files of similar length the system was still unable to differentiate between heartbeats and background noise in most cases. Kampouraki used support vector machines (SVMs) to classify ECG recordings [5]. However, the relatively clean ECG or simulated heart sounds of similar length are very dissimilar to real life data which often is of varying durations and with excessive background noise. Under such circumstances, the differentiation among different subtle heart symptoms can be extremely challenging. To cater to demands from such data, Liang chose Chebyshev type I low-pass filter combined with Shannon energy to attenuate noise and make the findings of low intensity sounds, namely heart beats, easier[6].

3 Data sets

Two datasets were provided for the challenge. Dataset A comprises data crowd-sourced from the general public via the iStethoscope Pro iPhone app. Dataset B comprises data collected from a clinical trial in hospitals using the digital stethoscope DigiScope [1].

The audio files are of varying lengths, between 1 second and 30 seconds (some have been clipped to reduce excessive noise and provide the salient fragment of the sound). Most information in heart sounds is contained in the low frequency components, with noise in the higher frequencies. It is common to apply a low-pass filter at 195 Hz. Fast Fourier transforms are also likely to provide useful information about volume and frequency over time. More domain-specific knowledge about the difference between the categories of sounds is provided below.

3.1 Normal Category

In the Normal category there are normal, healthy heart sounds. These may contain noise in the final second of the recording as the device is removed from the body. They may contain a variety of background noises (from traffic to radios). They may also contain occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. A normal heart sound has a clear “lub dub, lub dub” pattern, with the time from “lub” to “dub” shorter than the time from “dub” to the next “lub” (when the heart rate is less than 140 beats per minute). Note the temporal description of “lub” and “dub” locations over time in the following illustration:

…lub………dub……………lub………dub……………

In medicine we call the lub sound "S1" and the dub sound "S2". Most normal heart rates at rest will be between about 60 and 100 beats (‘lub dub’s) per minute. However, note that since the data may have been collected from children or adults in calm or excited states, the heart rates in the data may vary from 40 to 140 beats or higher per minute. Dataset B also contains noisy_normal data - normal data which includes a substantial amount of background noise or distortion.

3.2 Murmur Category

Heart murmurs sound as though there is a “whooshing, roaring, rumbling, or turbulent fluid” noise in one of two temporal locations: (1) between “lub” and “dub”, or (2) between “dub” and “lub”. They can be a symptom of many heart disorders, some serious. There will still be a “lub” and a “dub”. One of the things that confuses non-medically trained people is that murmurs happen between lub and dub or between dub and lub; not on lub and not on dub. Below, we illustrate with an asterisk* at the locations a murmur may appear:

…lub…****…dub……………lub…****…dub

or

…lub………dub…******…lub………dub…******…lub……dub

Dataset B also contains noisy_murmur data - murmur data which includes a substantial amount of background noise or distortion.

3.3 Extra Heart Sound Category (Dataset A)

Extra heart sounds can be identified because there is an additional sound, e.g. a “lub-lub dub” or a “lub dub-dub”. An extra heart sound may not be a sign of disease. However, in some situations it is an important sign of disease, which if detected early could help a person. The extra heart sound is important to be able to detect as it cannot be detected by ultrasound very well. Below, note the temporal description of the extra heart sounds:

…lub.lub………dub………………lub.lub………dub……………

or

…lub………dub.dub…………………lub………dub.dub

3.4 Artifact Category (Dataset A)

In the Artifact category there are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernable heart sounds, and thus little or no temporal periodicity at frequencies below 195 Hz. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again.

3.5 Extrasystole Category (Dataset B)

Extrasystole sounds may appear occasionally and can be identified because there is a heart sound that is out of rhythm involving extra or skipped heartbeats, e.g. a “lub-lub dub” or a “lub dub-dub”. (This is not the same as an extra heart sound as the event is not regularly occuring.) An extrasystole may not be a sign of disease. It can happen normally in an adult and can be very common in children. However, in some situations extrasystoles can be caused by heart diseases. If these diseases are detected earlier, then treatment is likely to be more effective. Below, note the temporal description of the extra heart sounds:

……lub………dub…………………lub…………………dub……………lub.lub ……………dub……

or

…lub………dub………………………lub…………………dub.dub………………… lub……………dub……

4 Heart Sound Segmentation

In Task 1 we try to produce a method which can locate the heartbeat and determine the sequence of S1 and S2 in the normal audio clips in Dataset A and Dataset B. Before segmentation, the signals are first de-noised using a combination of Short Time Fourier Transformation and wavelet [6][7].

Step1: Decompose the original signal using wavelet decomposition and reconstruct the approximations and part of the details. Re-filter the signal using the Spectrogram.

In order to identify S1 and S2 correctly, frequency band in which S1 and S2 concentrate should be used. Earlier studies [8] show that most information in heart sounds is contained in the low frequency components, with noise in the higher frequencies. Hence we first use the wavelet method to remove those details in higher frequency while preserve the main feature as approximation in lower band. Before decomposition, the original signal was down sampled by a factor of 10. Since the heart sound feature with the highest frequency is murmur which is up to 600Hz [8], the new sampling frequency 4410Hz is still more than two times higher. Thus no useful features of heart sounds are missed. After down-sampling, we adopt a forth-level Order Six Daubechies filter [6] to decompose the signal. Then we remove all the details in each level and use the approximation whose frequency is below 288Hz to reconstruct the signal. Finally we use Spectrogram to extract signal below 195Hz to further curb the noise.

Step 2: Find the peak location where the amplitude and slope exceed the selected threshold.

Even after pre-processing, the actual heart sound signal still has very complicated patterns with numerous small spikes that have little impact on diagnosis but may influence the location of S1 and S2. Hence we first use a "triangular smooth" [9] to smooth the signal. Then we calculate the derivatives of each point on the smoothed signal. After that, we mark those where the derivatives change from positive to negative. For each marked point, if the slope (the difference between the derivatives of the marked point and the one following it) and amplitude of it exceed the selected threshold, we use the polyfit function provided by Matlab to fit that point and its neighbourhood with parabola [10]. Peak-group is the number of points around the top part of the peak that are taken for measurement. By adjusting the size of the peak-group we can control the number of points around the peaks we want to use to fit the spikes, in other words, how detail we want to fit the signal. Finally, we pick those points closest to the peaks of the fitted parabolas as heartbeat. Amplitude threshold, slope threshold and peak-group all control peak sensitivity. Higher values will neglect smaller features.

Step3: Reject extra peaks

The ideal situation is where each spike that we select corresponds to one component of heartbeat, S1 or S2. However, in case one of S1 and S2 is too weak and to preserve the evidence of possible murmur and extra sound (extra-systole in Dataset B), we cannot set the threshold too harshly, which results in extra peaks. These extra peaks can provide useful details in classification but they are troublesome in identifying S1 and S2. To eliminate the extra peaks, we calculate the intervals between each adjacent peak. If two peaks appear within 50ms, which is the largest split normal sound interval, we choose the one with higher amplitude (in most cases, the real heart spike has larger energy than noise). If the interval of two peaks is larger than 50ms, we preserve both. By doing so, we preserve those with the highest energy and drop most of the possible split heart sounds.

Step3: Identify S1 and S2

After all the heart spikes have been recognized, we need to identify which of them are S1 and which are S2. Here our identification is mainly based on the interval. According to statistics, the systolic period is relatively shorter compared to the diastolic period. Hence we compare the mean of every other interval M1 and M2 and locate the larger one as the diastolic period and the shorter one systolic period. Unfortunately, this feature does not always work, especially in the case of children or adults with faster heart rates. When the heartbeat is above 120bpm, the difference between the length of diastolic and systolic period is extremely subtle.

5 Results for Challenge 1

In Table 1 and Table 2 we present the segmentation results for audio files in Normal groups in Dataset A and Dataset B, respectively. For precision, the error here is measured in samples. The unit of heartbeat is beat per minute.

Table 1 Results for Dataset A

Dataset A	Heart beat	Avg Err
201101070538	11.5	43380.56
201101151127	10.5	211373.76
201102081152	5.5	113710
201102201230	11.5	17118.69
201102270940	1	1445798.5
201103101140	9	71534.77
201103140135	5	314214.6
201103170121	5.5	299207.54
201104122156	2.5	509360.4
201106151236	5	368680

As can be seen, the results for Dataset B are much better than those for of Dataset A. The total error of Dataset B is 75569.78488 while that of Dataset A is 3394378.846. This is likely to be because data in Dataset B are collected in hospitals by experts under quieter conditions while those in Dataset A are produced by non-experts in highly variable conditions.

Table 2 Results for Dataset B

Dataset B	Heart beat	AveErr
103_1305031931979_B	12.5	50.32
103_1305031931979_D2	10.5	1013.
106_1306776721273_B1	4	58.75
106_1306776721273_C2	3	79.33
106_1306776721273_D1	4	1723
106_1306776721273_D2	8	4079.31
107_1305654946865_C1	8	2845.62
126_1306777102824_B	6.5	12070.76
126_1306777102824_C	3.5	11024.85
133_1306759619127_A	4.5	1629.88
134_1306428161797_C2	2.5	74.8
137_1306764999211_C	15	72.93
140_1306519735121_B	13	8556.38
146_1306778707532_B	19	4813.89
146_1306778707532_D3	3	37.33
147_1306523973811_A	5	4242.5
148_1306768801551_D2	9.5	3393.42
151_1306779785624_D	4.5	320.77
154_1306935608852_B1	4.5	62.66
159_1307018640315_B1	7	3558
159_1307018640315_B2	3	60.33
167_1307111318050_A	13.5	2147.88
167_1307111318050_C	5.5	3416.72
172_1307971284351_B1	3.5	63.71
175_1307987962616_B1	2.5	28
175_1307987962616_D	10.5	7260.52
179_1307990076841_B	16.5	98.84
181_1308052613891_D	3.5	1405.71
184_1308073010307_D	26.5	83.64
190_1308076920011_D	4.5	1296.11

6 Heart Sound Classification

Task 2 aims to produce a method that can classify the heartbeat audio into Normal, Murmur, Extra Heart Sound for Dataset A and Normal, Murmur and Extra-systole for Dataset B. Files Aunlabelledtest and Bunlabelledtest are provided to evaluate the performance of the method.

To classify the heart sound, we mainly depend on the number of heartbeat and features of systole and diastole period. In Figure 1 we present the flow chart of classification for Dataset A. Dataset B follows a similar procedure. L1 is the length of the peak sequence before extra-peak-rejection and L_s is the length of finally selected peak sequence after rejection. M1 and M2 are the mean of the systole and diastole period. N denotes the number of heartbeats per minute calculated from dividing 60 by the sum of M1 and M2. Std1 and Std2 are the standard deviation of diastole period and systolic period. Std12, Std 22, Std13 and Std23 are the new standard deviation of diastole and systolic period after dropping the smallest interval and the longest interval among the finally selected S1 and S2, respectively.

After segmentation, N is calculated for each sound clip first. If N is within the range from 30 to 140 beats per minute, the clip goes into the next level of judgement. If not, it is labelled as artefact. Though in most cases the normal heart rates in the data should vary from 40 to 140 beats per minute, we find several clips with heart rates between 30 and 40. Hence we slightly modify the boundary. The next level compares the length of the signal before and after extra-peak-rejection. If the L1 is more than 3.3 times of L_s, which means other than S1 and S2 the signal contains a considerable number of spikes, we categorize those clips into the Murmur group considering: 1. Most of the noise has already been filtered; 2. The process of curve-fitting would further filter those sudden spikes highly likely to be un-filtered noise, which indicates most of the leaving spikes are the reflection of heart condition. The last level focuses on the mean and standard deviation systole and diastole period. If either standard deviation is larger than its corresponding mean, or either standard deviation drops obviously after removing the smallest interval or largest interval, then the signal would be classified into Extra sound group. If not, then we label them as normal heartbeat.

7 Results for Challenge 2

We evaluate our method based on three metrics calculated from the TP (true positives), FP (false positives), TN (true negatives) and FN (false negatives). They are precision for each class, the Youden's Index, the F-score (for Dataset A) and the Discriminant Power (DP) (for Dataset B) [11]. Precision estimates the percentage of correctly classified samples in result of each class. Youden's Index has traditionally been used to compare diagnostic abilities of two tests [11]. Here it evaluates the algorithm's ability to tell artifact sound clips from non-artifact ones. For Dataset B the metrics assess the algorithm's ability to differentiate problematic heartbeats (including murmur and extra-systole) from normal ones. F-score considers both precision and specificity for the artefact group [11]. Here F-score is tuned to favour slightly more the specificity. In other words, we would rather the algorithm mis-classifies non-artefact sound clip as artefact than having one that may categorize artefact sound into humans heartbeat. DP provides another insight into how well the algorithm distinguishes between the algorithm's ability to tell artefact sound clips from non-artefact ones. The algorithm is a poor discriminator if DP<1, limited if DP<2, fair if DP<3, good in other cases.

Tables 3 and 4 summarise the results for the challenge.

We can see from tables 3 and 4 that in Task 2 our algorithm stills performs much better in Dataset B, with the precision of normal group increasing sharply. Part of the reason is the influence of the Task 1. It is clear that the real-world noise present in the background of audio files for Dataset A present tremendous challenges for segmentation and classification, despite the fact that the quality of the audio is superior. In both datasets the algorithm does not perform well on recognizing the extra sound group. In fact, in the test on the training group, most of the samples in extra sound group (Extra systole in Dataset B) are classified into the normal group. This is likely to be because most extra sound clips only have one or two abnormal spikes while the rest of them are normal ones.

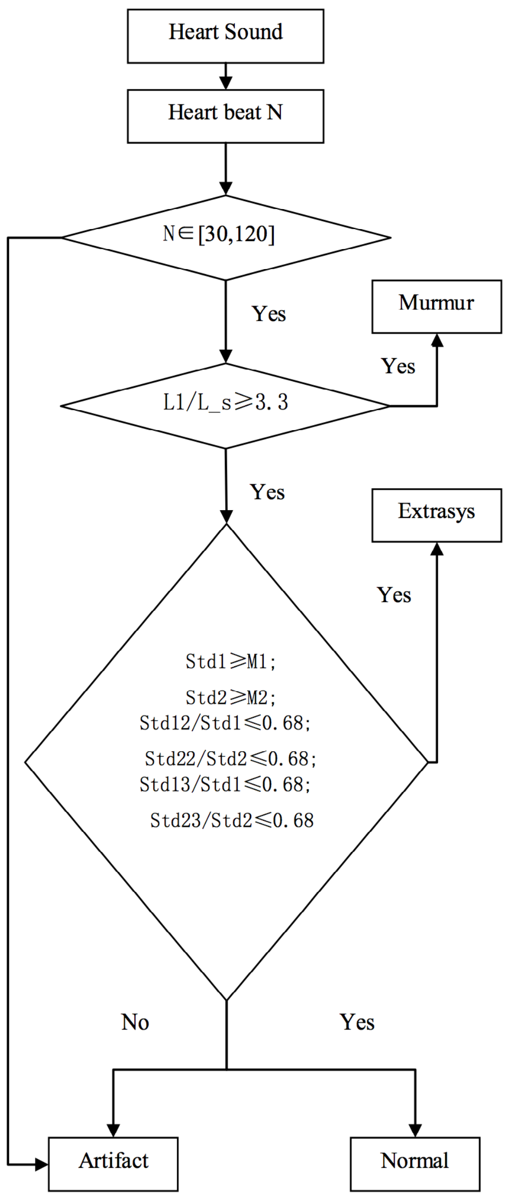

Figure 1 Flow chart of Classification for Data set A

Table 3 Results of classification for Dataset A

Dataset A	
Precision of Normal	45.83%
Precision of Murmur	31.25%
Precision of Extras	11.27%
Precision of Artifact	58.33%
Artifact Sensitivity	43.75%
Artifact Specificity	44.44%
Youden Index of Artifact	-0.0902
F-score	0.1396
Total Precision	1.4668

Table 4 Results of classification for Dataset B

Dataset B	
Precision of Normal	77.67%
Precision of Murmur	36.99%
Precision of Extrasound	16.67%
Heart problem Sensitivity	50.85%
Heart problem Specificity	58.82%
Youden Index of Heart problem	0.0967
Discriminant Power	0.0935
Total Precision	1.3132

8 Conclusions

In this paper, we present our approach for a segmentation and classification method which would be less sensitive to ambient noises and recording locations compared to existing methods, and uses the heart sound signal as the only source. We found that past algorithms that showed good performance on ECG could not properly handle real life data. More specifically, the more specialised the algorithm, the more unstable when it faces real-world heartbeat recordings. To address these issues, in this work we create an improved de-noise algorithm by combining wavelet and spectrogram. Amplitude and slop thresholds are used to control the sensitivity of peak finding. We then realign the peaks by exploiting the interval features. In the classification part, we exploit domain knowledge and compare the features of heartbeat before and after dropping out extra peaks and those before and after dropping out the smallest interval. By doing so we try to minimize the possible effect of excessive noise and realize better robustness. We find in both tasks, the method works better for Dataset B, which contains data with less noise.

Acknowledgements

The Classifying Heart Sounds Challenge is sponsored by the PASCAL Network of Excellence.

References

[1] The PASCAL Classifying Heart Sounds Challenge 2011, Bentley P. and Nordehn G. and Coimbra M. and

Mannor S. Available: http://www.peterjbentley.com/heartchallenge/index.html

[2] M. W. Groch, J. R. Domnanovich, and W. D.Erwin, “A New Heart-Sounds Gating Devices for Medical Imaging”, IEEE Trans. On Biomedical Engineering, Vol.39, No.3, pp.307-310, March 1992

[3] S. L. Strunic, F. Rios-Gutierrez, R. Alba-Flores, G. Nordehn, S. Burns, “Detection and Classification of Cardiac Murmurs using Segmentation Techniques and Artificial NeuralNetworks”, IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2007, pp.397-404

[4] G.Karraz, G. Magenes, “Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework”, IEEE, EMBS Annual International Conference, New York City, USA, Aug 30-Sept 3, 2006

[5]Argyro Kampouraki, George Manis, and Christophoros Nikou, “Heartbeat Time Series Classification with Support Vector Machines”, IEEE Transformation on Information Technology in Biomedicine, Vol. 13, No.4, July 2009

[6] H. Liang, S. Lukkarinen, I. Hartimo, “Heart Sound Segmentation Algorithm Based on Heart Sound Envelogram”, Computers in Cardiology 1997, September, Lund, Sweden

[7] Cheng Y., “Music Database Retrieval Based on Spectral Similarity”, International Symposium on Music Information Retrieval (ISMIR) 2001, Bloomington, USA, October 2001

[8] R. M. Rangayyan and R. J. Lehner, “Phonocardiogram Signal Analysis: A Review”, CRC Critical Reviews in Biomedical Engineering, Vol.15, Issue 3, pp.211-236, 1988

[9]Tom O'Haver, “Intro. To Signal Processing: Smoothing”, Available: http://terpconnect.umd.edu/~toh/spectrum/Smoothing.html

[10] Polyfit, Available:Http://www.mathworks.co.uk/help/techdoc/ref/polyfit.html

[11] Peter A. Flach, “The Geometry of ROC Space: Undstanding Machine Learning Metrics through ROC Isometrics”, Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003, pp.226-233