

Using Spectral Acoustic Features to Identify Abnormal Heart Sounds

Nicholas E. Singh-Miller^{1,2}, Natasha Singh-Miller

¹ Naval Medical Center Portsmouth, Portsmouth, VA, USA

² Uniformed Services University of Health Sciences, Bethesda, MD, USA

Abstract

Using the Physionet challenge database we aim to determine whether a heart sound recording, a phonocardiogram (PCG), corresponds to a “normal” or “abnormal” physiological state. Our goal is to augment the information available to a physician during auscultation of a patient’s heart, ultimately assisting with clinical decision making. To that end, we first produce spectral features of the PCG, for varying windows and frequency bands. We use the resulting spectral information to identify a variety of features based on means, variance, and activity at different frequency bands. We find that much of the information corresponding to abnormalities is captured in these features, with particular good performance on murmurs. Finally, we build a discriminative model, specifically a random forest regressor, to classify new samples based on the aforementioned features. Our final performance on the challenge data received a combined score of 81%.

1. Introduction

Auscultation of the chest is an element of the medical physical exam that has remained largely unchanged since the times of the French physician Laënnac in 1816 [1]. The normal sounds of the heart produce a periodic signal in frequency ranges that are audible to the human ear with the aid of a stethoscope. Due to this accessibility and the importance of cardiac auscultation for screening and identifying pathological cardiac conditions, attempts to augment the skill of auscultation with signal processing have been pursued for half a century.

Past methods to augment and/or automate heart sound auscultation are numerous, but in general, the approach has been taken in three steps. First the PCG is segmented into the typical physiological events S1, systole, S2, diastole, etc. This step has been approached with success with many different methods from established signal processing techniques to machine learning; overviews and reviews of these methods can be found in Liu *et al.* [2], Emmanuel [3], and in Choi, *et al.* [4]. Second, feature extraction on the segmented PCG is performed. In this

step there is a huge amount of variability in features used in the literature, however they are typically from one of four domains: time, frequency, statistical, and a combined time-frequency [2, 5]. Relevant to our approach numerous previous works have used the energy profile at different frequency bands as features; noting that pathological states often have persistent signals at frequencies higher than normal heart sounds [2]. The third step is typically classification; relevant to our methodology, there are past approaches that use machine learning for classification based on the aforementioned features, again summary reviews can be found [2, 5]. Therein, it is pointed out the most prevalent methodologies in the existing literature for machine learning based classification are artificial neural networks (ANN) and support vector machines (SVM); there are numerous other methods in use, however the goals of any particular study are often different (such as systolic murmur detection, valvular defect detection, etc.), and the machine learning classification method of choice is often tailored to meet these specific goals [5].

2. Methodology

In this paper we bypass segmentation as it is described in the above references, instead opting for a methodology that uses a combination of unsupervised segmentation and clustering of spectral data to build a subsequent discriminative model.

The data used for this work is from a public database provided by Physionet, and is described in detail in Liu *et al.* [2]. The data consists of more the 3,000 heart sound recordings that were obtained at numerous clinics around the world from patients of all ages. One express goal of the challenge is to retain recordings that are corrupted with noise (bowel sounds, speech, etc.) with the intent that more robust algorithms can be developed for real world situations (such as outpatient home visits, etc.). Each recording is labeled either “normal” or “abnormal” based on whether the recording was made from a healthy individual or an individual with a confirmed cardiac condition, respectively. Using the labeled data as a training set, the object of the challenge is to accurately classify an unlabeled test set of

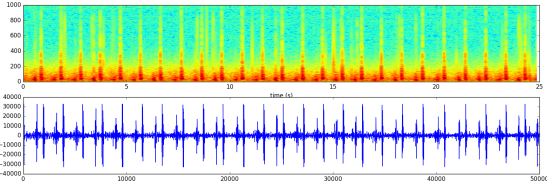


Figure 1. Spectrogram (top) and audiogram (bottom) for file a0118 in the training data set that was labeled “normal”.

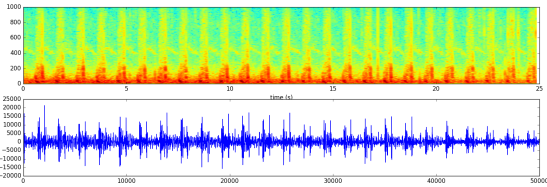


Figure 2. Spectrogram (top) and audiogram (bottom) for file a0343 in the training data set that was labeled “abnormal”.

heart sounds. Scoring, for the purpose of the challenge, is based on a modified sensitivity (Se) and specificity (Sp) that are weighted based on the quality of the recording, the full details can be found in [2].

2.1. Spectral Features

Each audiogram A , is a variable length sequence of acoustic measurements, $A = [a_1, a_2, \dots, a_T]$, where $a_i \in \mathcal{R}$ represents the acoustic measurement at time i . In order to visualize the trends in frequency versus time within each audiogram, we compute the corresponding spectrogram using the `signal.spectrogram` method of the `scipy` package [6]. In our approach we use Gaussian windows with window lengths of 15 and 75. Computation of the spectrogram leaves us with a spectral feature representation $f(A) = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T]$ where $\mathbf{f} \in \mathcal{R}^L$ is a real-valued vector of length 1001. Sample spectrograms for both a normal and abnormal heart sound recording are shown in Figures 1 and 2, respectively.

2.2. Feature Representation

For each acoustic recording, we start with its spectral feature representation $f(A)$. We normalize by the maximum value along each dimension of the sequence of feature vectors, which corresponds to a particular frequency, and reduce each vector \mathbf{f} down to one hundred features by averaging along equally spaced intervals of the frequency spectrum (from 0-1000 Hz). We then perform clustering of the vectors, by successively projecting them down two

dimensions using principal components analysis, and then using k-means clustering. On each iteration we bisect the largest remaining cluster, and retain five total clusters. We eliminate clusters that are infrequent, indicating they may correspond to noise events. A cluster is considered infrequent if it has fewer than three ‘spans’, where a ‘span’ is a maximal sequence of time where all consecutive samples are classified within that cluster. Figure 3 shows the clustering derived for an example acoustic recording.

We then calculate the following features:

- the mean values of all spectral features in the bottom 25 intervals (0-250Hz), next 25 intervals (250-500Hz), and top half of the intervals (500-1000Hz), for both window sizes
- the variance values of all spectral features along the same intervals for both window sizes
- the mean values of all non-normalized spectral features along the same intervals for window length 75
- the variance values of all non-normalized spectral features along the same intervals for window length 75
- the mean values of all spectral features in the bottom 25 intervals (0-250Hz) with data normalized by maximum value at every point in time for window length 75
- the average number of spans of the first cluster per second for both window lengths

We also calculate a set of features designed to model activity at different frequencies. We look at specific frequency intervals (2,5,10,20,30,40,50,75 of 100 total intervals) throughout the entire signal. We take the values of $f(A)$ along a particular frequency interval r , call it \mathbf{y}_r , sort them and assign them to vector \mathbf{x} of the same length with values equally spaced from 0 to 1. We then fit the resulting (x, y) pairs from $(\mathbf{x}, \mathbf{y}_r)$ to a logistic function $y = \frac{L}{1 - \exp^{-k*(x-x_0)}}$. Figure 4 shows an example of a logistic function fit for a data sample. We use the estimated parameters L , k , and x_0 as features for the identified intervals for both window lengths.

2.3. Discriminative Classification

We use our feature representation to train a discriminative learning model, specifically a random forest regressor, in order to predict the label of a new test sample. We perform feature reduction by selecting the 25 features with the highest feature importance, as determined by the trained model and retrain with just those 25 features. We utilize 500 estimators within the random forest regressor and classify examples as positive or negative by adding 0.4 to the predicted regression value and using the sign of the resulting value. This threshold was determined by optimizing over cross-validation on the training set. Ten-fold cross-validation over the training set results in a score of 84.8%.

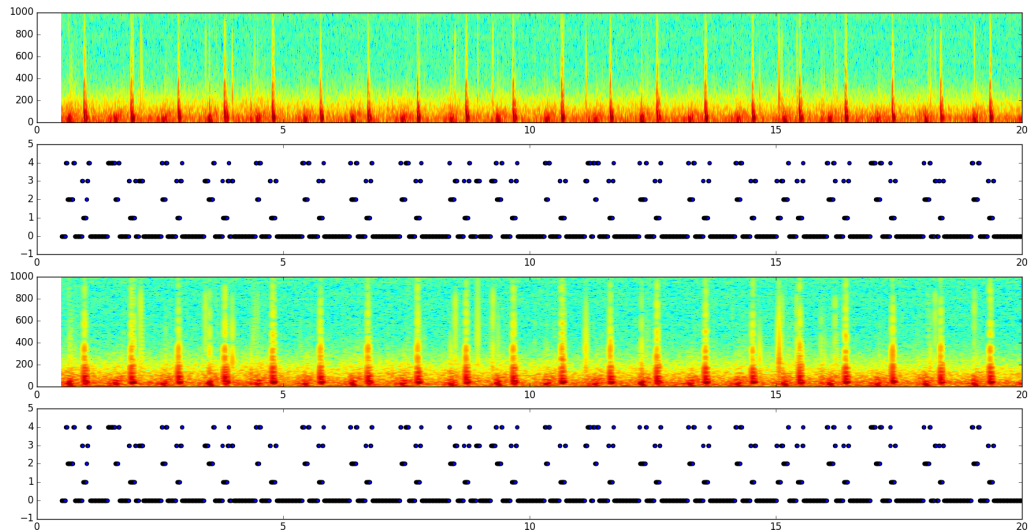


Figure 3. Spectrograms and associated clusterings for our two selected window lengths for file a0118.

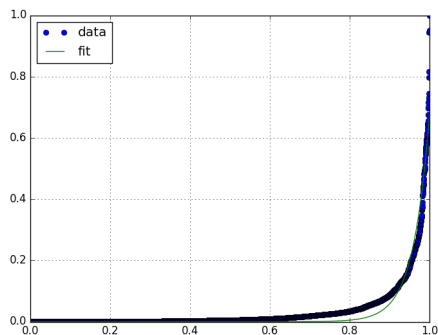


Figure 4. Logistic function fit for sixth frequency interval (50Hz-60Hz) for file a0014.

3. Results and Discussion

Our final model received an overall score of 81% which correlates well to our cross validation, see Table 3. It should be noted that we do not label any of the files as “unsure” in our model. We perform well against the binary linear regression (BLR) benchmark provided in [2], which uses 20 features derived from the segmentation of the PCG described in the paper. Our approach does not explicitly attempt to model the different parts of a heart sound (S1, S2, etc.) or attempt to segment the signal explicitly, and on cross-validation with only one feature (power in high frequency ranges) we are able to achieve scores in the 60s.

Beyond the labels normal and abnormal, the data set

Table 1. Result of the challenge submission, with a cross-validation score, and the BLR benchmark.

	Se	Sp	$\frac{Se+Sp}{2}$
Challenge	0.76	0.87	0.81
Cross Validation	0.81	0.89	0.85
BLR Benchmark	0.62	0.70	0.66

also contains corresponding labels of specific diagnoses for each recording. These labels were not included in the training of our model. This labeling comprises a list that is not comprehensive of all the possible audible pathologies, but they can be grouped into broader categories. For instance, murmurs are usually associated with mitral valve prolapse (MVP), aortic stenosis (AS), and mitral regurgitation, while there is not typically an auscultatory finding associated specifically with coronary artery disease (CAD). With that in mind, we can clearly see that our model has more difficulty with certain abnormalities, Table 3.

Here we see that we perform well on murmur related pathologies. This result was somewhat expected, given that the spectral signature of a murmur tends to be present at high amplitude throughout the studied frequency range, and the features in our model highlight this phenomenon well, particularly in the high frequency ranges. A good example of this can be seen Figure 2, where a high amplitude band is seen occupying the entire frequency range between S1 and S2 for every beat (likely a holosystolic murmur). By contrast, we can see the lack of high frequency

involvement of the “normal” spectrogram in Figure 1.

Table 2. Comparisons of incorrectly labeled files during cross-validation based on the diagnoses provided.

Pathology	No. incorrect	Total No.	%
Normal	394	2488	16
MVP	10	134	7.5
Aortic Disease and AS	2	29	6.9
Benign	3	118	2.5
Mitral Regurgitation	0	12	0.0
CAD	65	294	22
Other Path. and MPC	19	97	20

We see that coronary artery disease (CAD) and the “other pathologic” findings category are more difficult for our model. This poorer performance can possibly be attributed to the timing and frequency of these phenomenon excluding them from our feature selection. For instance, while CAD is not usually heard while auscultating, it is a change in the blood flow through the coronary arteries and should logically cause turbulence and a subsequent murmur-like sound. Recent studies have investigated the clinical viability of seeking out this murmur, and have shown that with a sensitive recording stethoscope a diastolic murmur that corresponds to the extent of arterial disease can be found in frequencies under 500Hz with the bulk of the signal occurring under 150Hz [7]. This sound will be many orders of magnitude softer than the other normal heart sounds and will effectively be “buried” in this lower frequency range. However, unsupervised clustering in this low frequency range with logarithmic scaling of the power could potentially provide some finer detail resolution.

The catch-all category of “other pathologic” and MPC (misc. pathologic condition) likely includes all of the non-murmur pathologies such as heart sound splitting and extra hearts sounds such as S3 and S4. Taking into consideration the frequency ranges of sounds such as S3 and S4 (illustrated in Fig 2 in [2]) we expect poor performance similar to CAD due to the nature of our feature selection. However, splitting of S1 or S2 should occur in the same frequency ranges as the normal S1 and S2. Thus it is less likely that our feature selection misses these in the same manner as CAD, but rather the short time duration of the signal does not have a large enough effect on the features meant to capture activity at different frequency bands. Again, unsupervised clustering may be able to provide features that are more amenable to detecting splits in S1 and S2; an example of how this clustering can straight forwardly pick out S1 and S2 can be seen in Figure 3.

In conclusion, we have built a model for determining normal versus abnormal heart sounds that relies heavily

on the spectral features of the PCG to derive features for the discriminative model. Despite not segmenting the PCG into typical physiologic heart sound landmarks, we achieve modestly good results.

Acknowledgements

Thank you to Dr. Michael Mattingly, Dr. Michael Cunningham, and Dr. Virginia Randall for fruitful conversations and guidance.

I am a military service member. This work was prepared as part of my official duties. Title 17 U.S.C. 105 provides that “Copyright protection under this title is not available for any work of the United States Government.” Title 17 U.S.C. 101 defines a United States Government work as a work prepared by a military service member or employee of the United States Government as part of that person’s official duties.

The views expressed in this article are those of the author(s) and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the United States Government.

References

- [1] Seaton A. Alas, poor laënnec! *QJM* 2011;104(3):275–277.
- [2] 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. *Physiological Measurement* 2016;37(9).
- [3] Emmanuel BS. A review of signal processing techniques for heart sound analysis in clinical diagnosis. *Journal of Medical Engineering Technology* ;36(6):303 – 307.
- [4] Choi S, Jiang Z. Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Systems with Applications* 2008;34(2):1056 – 1069.
- [5] Leng S, Tan RS, Chai KTC, Wang C, Ghista D, Zhong L. The electronic stethoscope. *BioMedical Engineering OnLine* 2015;14(1):1–37.
- [6] Jones E, Oliphant T, Peterson P, et al. *SciPy: Open source scientific tools for Python*, 2001–. URL <http://www.scipy.org/>.
- [7] Schmidt SE, Holst-Hansen C, Hansen J, Toft E, Struijk JJ. Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering* 2015; 62(11):2611–2619.

Address for correspondence:

Nicholas E. Singh-Miller, MD, PhD
Capt, USAF, MC

620 John Paul Jones Circle, Portsmouth, VA 23708-2197
nicholas.e.singhmiller.mil@mail.mil