

Power Spectrum Analysis for Classification of Heart Sound Recording

Soo-Kng Teo¹, Bo Yang¹, Ling Feng¹, Yi Su¹

¹Institute of High Performance Computing, A*STAR, Singapore

Abstract

This study assessed the feasibility of using power spectrum analysis to compute additional features from heart sound recordings that can be used for normal / abnormal classification in the PhysioNet/Computing in Cardiology Challenge 2016. Our approach relies on the segmentation of the heart sound recording into the 4 distinct Fundamental Heart Sounds (FHS) states. Based on the FHS states, the heart sound recording is divided into multiple segments - each segment corresponds to 1 FHS state in 1 cardiac cycle. A fast Fourier transform is then performed on each of these segments. For each FHS state, we compute the ratio between the sum of the powers in the N highest peaks in the power spectrum to that of the entire power spectrum. The rationale is that the most dominant frequencies in the heart sound recording can potentially contain relevant information useful for classification. These additional features are subsequently combined with the features computed from the intervals of the 4 distinct FHS states. This enlarged set of features is used to train a feedforward neural network with 1 hidden layer for heart sound clarification. The scores of our neural network on a random subset of the test data are as follows: Sensitivity = 0.747; Specificity = 0.788; Overall = 0.767.

1. Introduction

The objective of this paper is to assess the feasibility of using power spectrum analysis to compute additional features from heart sound recording that can be used for normal / abnormal heart sound classification. Power spectrum analysis is a well-established method for computing the distribution of the dominant frequencies in signal processing. An uncorrupted (clean) sinusoidal signal will be represented by one single peak in the power spectrum corresponding to the angular frequency of the signal. When this sinusoidal signal is corrupted by white noise (noise with zero mean) of low amplitude compared to the original signal amplitude, the power spectrum will be represented by multiple peaks. It is still possible to identify the frequency of the original signal by searching for the most dominant frequency in the power spectrum.

Based on this reasoning, it is plausible that power spectrum analysis can be used to identify additional features for heart sound classification. Intuitively, we will expect that the heart sound recording of a normal healthy individual to be dominated by a small subset of frequencies that correspond to the anatomical contraction and relaxation of the heart. This is because the contraction and relaxation are activated by regular and periodic electrical activities. In contrast, the heart sound recordings for patients with cardiovascular diseases will be corrupted with multiple sources of noise corresponding to the anatomical / structural heart defects [1].

2. Methods

Our approach relies on the ratio between the sum of the powers in the N highest peaks in the power spectrum to that of the entire power spectrum. The rationale is that the most dominant frequencies in the heart sound recording can potentially contain relevant information useful for classification. The detailed step for our approach is as follows:

(i) The heart sound recording is segmented into 4 distinct Fundamental Heart Sounds (FHS) states, namely S1, systole, S2 and diastole. This segmentation is performed using the algorithm of Springer *et al.* [2].

(ii) Based on the FHS states, the heart sound recording is divided into multiple segments – each cardiac cycle in the recording will be represented by 4 individual segments corresponding to the 4 FHS states. All segments corresponding to the same FHS state are grouped into the same set.

(iii) For each set, a fast Fourier transform is performed on each of these individual segments (see Figure 1). The power spectrum from the fast Fourier transform is then sorted in descending order.

(iv) We compute the sums S_N , where S_N is the sum of the N highest peaks in the power spectrum. Ratios of S_N to the sum of the entire power spectrum are also computed. The mean and standard deviation (SD) of these ratios for each set are then used as additional features for classification. In this way, we have $2N * 4$ additional features ($2N$ features for each FHS states) for each heart sound recording. The number of sums, N , can be optimized to increase the accuracy of the training

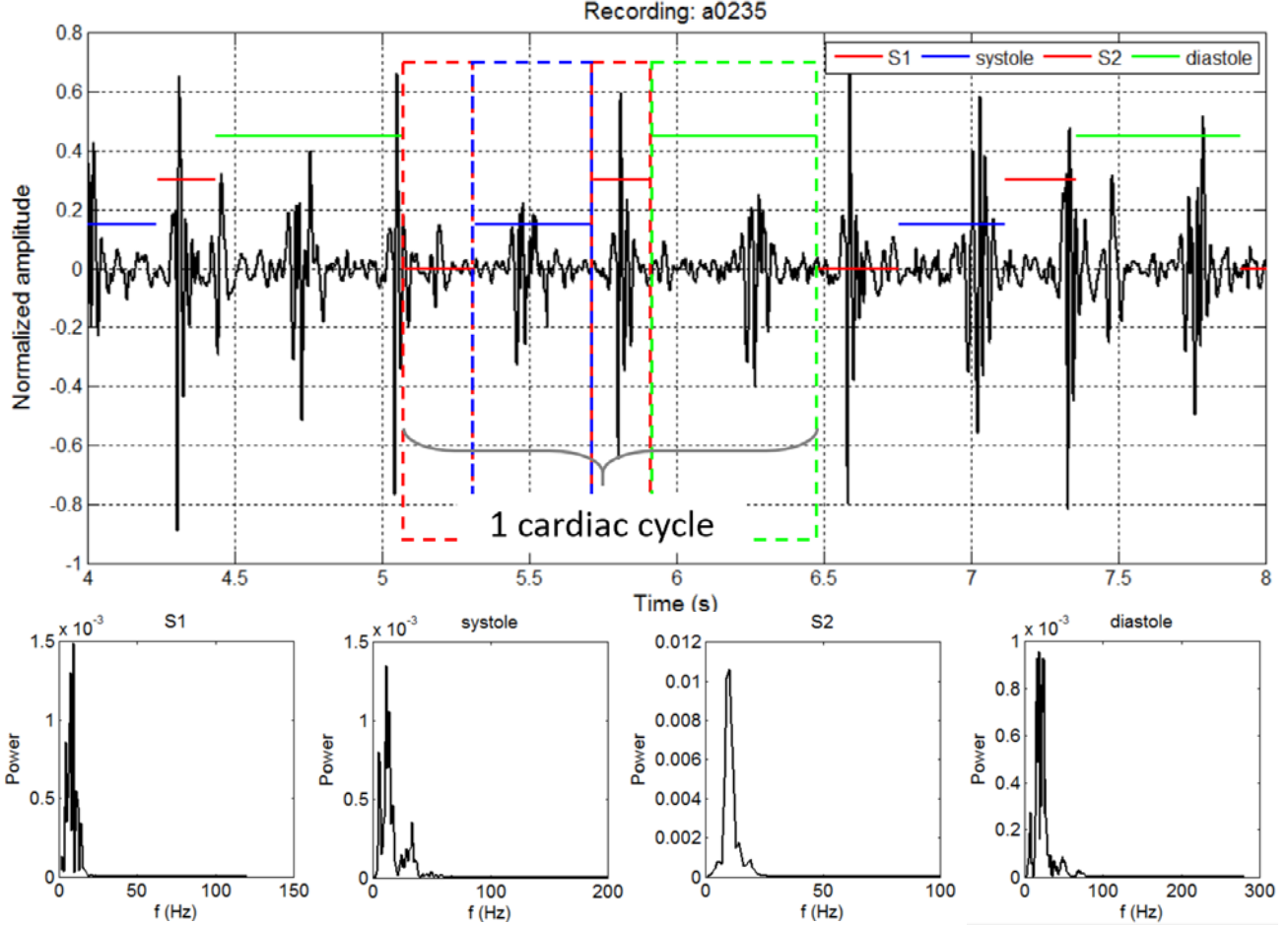


Figure 1. Illustration of the power spectrum analysis for a sample heart sound recording. Top panel: Overlaid of the heart sound recording with the FHS state labels (S1, systole, S2 and diastole) generated using the segmentation algorithm of Springer *et al.* [2]. For each cardiac cycle, the heart sound recording is divided into 4 segments (example of such segments in 1 typical cardiac cycle are represented by the colored rectangles above) corresponding to the 4 FHS. Bottom panel: Power spectrum analysis for 1 typical cardiac cycle. For each segment, we perform a fast Fourier transform to compute the power spectrum. These power spectrums are then used to compute the corresponding sums S_N .

prediction. For this paper, we used two sums; $N = 5$ and $N = 10$ to compute the additional features, resulting in 16 additional features. The definition of these 16 features are described in Section 2.1 below.

(iv) A balanced dataset is selected from all training sets provided in this challenge. This is necessary as the number of abnormal and normal heart sound recordings in the training sets are unbalanced. Details of the distributions of the normal and abnormal heart sound recordings in the training sets are described in Liu *et al.* [3]. The power spectrum analysis and segmentation interval analysis are performed on this balanced dataset. Details of the segmentation interval analysis and the associated features are described in [3].

(v) A feedforward neural network with regularization is trained using the combined features from both the power spectrum analysis and the segmentation interval

analysis to perform normal and abnormal heart sound recording classification. The neural network used for our classification have the following architecture: Input layer: 36 nodes, Hidden Layer: 10 nodes and Output layer: 2 nodes. The training of the neural network is implemented as an optimization problem in MATLAB [4].

2.1. Power spectrum feature definition

There are 16 features that are computed from the power spectrum analysis described in the section above. Let R1 be defined as the ratio of S_5 (S_5 is the sum of the 5 highest peaks in the power spectrum) to the sum of the entire power spectrum for a particular segment. Similarly, let R2 be defined as the ratio of S_{10} to the sum of the entire power spectrum for a particular segment. The detailed definition of the 16 features are as

follows:

1. m_{S1_r1} : mean value of R1 ratios for all S1 segments
2. sd_{S1_r1} : SD of R1 ratios for all S1 segments
3. m_{S1_r2} : mean value of R2 ratios for all S1 segments
4. sd_{S1_r2} : SD of R2 ratios for all S1 segments
5. m_{Sys_r1} : mean value of R1 ratios for all systolic segments
6. sd_{Sys_r1} : SD of R1 ratios for all systolic segments
7. m_{Sys_r2} : mean value of R2 ratios for all systole segments
8. sd_{Sys_r2} : SD of R2 ratios for all systolic segments
9. m_{S2_r1} : mean value of R1 ratios for all S2 segments
10. sd_{S2_r1} : SD of R1 ratios for all S2 segments
11. m_{S2_r2} : mean value of R2 ratios for all S2 segments
12. sd_{S2_r2} : SD of R2 ratios for all S2 segments
13. m_{Dia_r1} : mean value of R1 ratios for all diastolic segments
14. sd_{Dia_r1} : SD of R1 ratios for all diastolic segments
15. m_{Dia_r2} : mean value of R2 ratios for all diastolic segments
16. sd_{Dia_r2} : SD of R2 ratios for all diastolic segments

Table 1 shows the average values for all 16 features for normal and abnormal heart sound recordings on the selected balanced dataset. A non-parametric test (Wilcoxon rank sum test) for features without normal distribution is performed to check for statistical differences between the two groups. For features with normal distribution, a parametric Student t-test is performed. In summary, 11 features are statistically different between the normal and abnormal groups.

2.2. Regularization and validation

To prevent the neural network from over-fitting to the training data, a regularization parameter (λ) is introduced. This regularization parameter modifies the objective function in the optimization of the neural network by penalizing large weights that are assigned to the nodes of the network. In order to select the optimal value for λ , a 10-fold cross validation is performed on the balanced dataset for the following values of λ : 0.01, 0.03, 0.05, 0.08, 0.1, 0.3, 0.5, 0.8, 1.0, 1.5, 2.0, 3.0, 3.5, 4.0, 4.5 and 5. The mean of the overall score across the 10-fold on the test set is used as a criterion to select λ (see Figure 2 and Table 2).

3. Results and discussion

The 10-fold cross validation results for our neural network classification (for $\lambda = 1.5$) on the balanced dataset are shown in Table 2. We obtained a score of 0.90 ± 0.01 (0.73 ± 0.03) on the training (test) set averaged across the 10-fold. For the testing dataset on the

challenge website, our approach achieved the following score for a random subset: **Sensitivity = 0.747**, **Specificity = 0.788** and **Overall = 0.767**

Potential approaches to further optimize the performance of our neural network classification includes the following: (i) increasing the number of sums (S_M) used for computing the power spectrum features, (ii) modifying the values of N to either include or decrease the number of peaks used to compute the power spectrum features, and (iii) removing features that are not statistically different across the abnormal and normal heart sound recordings.

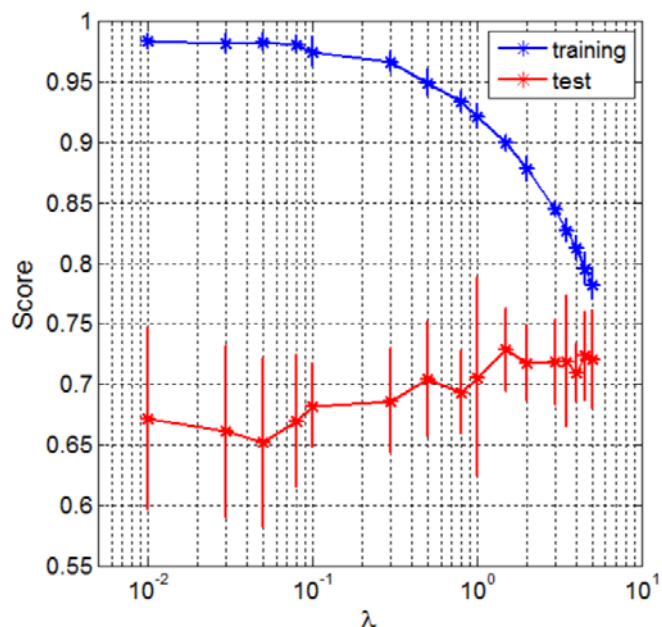


Figure 2. Selection of regularization parameter (λ) for the neural network. The overall score (mean \pm standard deviation) averaged across the 10-fold is plotted against the values of λ in a semi-log plot. As λ increases from 0.01 to 1.5, the score on the training (test) set decreases (increases) as over-fitting is penalized. For values of λ greater than 1.5, the score on the test set remains fairly constant while the score on the training set continue decreasing. Hence, we choose $\lambda = 1.5$ as the optimal regularization parameter for our neural network.

4. Limitations

The accuracy of our neural network classification is heavily dependent on the accuracy of the FHS state segmentation algorithm. This is because the power spectrum analysis requires the division of the heart sound recording into the 4 distinct FHS state. It is possible that the default parameters in the algorithm of Springer *et al.* [2] may not be optimized for certain recordings. Further optimization of the parameters used for FHS segmentation may potentially improve the accuracy of our approach.

5. Conclusion

We have shown that the additional features derived from power spectrum analysis can be used successfully for the classification of heart sound recordings. Our approach relies on the fast Fourier transform of segments corresponding to the 4 FHS states in the heart sound recording to compute these additional features. These additional features are combined with the features computed from the segmentation interval analysis and used to train a feedforward neural network with regularization. This neural network is then used for classification of heart sound recording.

References

- [1] El-Segaier M, Pesonen E, Lukkarinen S, Peters K, Sörmö L, Sepponen R. Detection of cardiac pathology: time

- intervals and spectral analysis. *Acta Paediatr.* 2007; 96(7):1036-42.
- [2] Springer DB, Tarassenko L, Clifford GD. Logistic Regression-HSMM-Based Heart Sound Segmentation. *IEEE Trans Biomed Eng.* 2016; 63(4):822-32.
- [3] Liu C Y, Springer D B, Li Q, Moody B, Juan R A, Chorro F J, Castells F, Riog J M, Silva I, Johnson A E W, Syed Z, Schmidt S E, Panadaniil C D, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab M R, Sameni R, Mark R G and Clifford G D 2016 An open access database for the evaluation of heart sound algorithms *Physiol Meas* 37 (in press).
- [4] MATLAB Release 2013a, The MathWorks, Inc., Natick, Massachusetts, United States.

Address for correspondence.

Soo-Kng Teo.

1 Fusionopolis Way, #16-16 Connexis, Singapore 138632

teosk@ihpc.a-star.edu.sg

Table 1: Comparison of the features derived from power spectrum analysis between the abnormal and normal heart sound recording from the balanced dataset. Values are presented as mean \pm standard deviation.

	Features	Abnormal	Normal	P-value
S1	m_S1_r1	0.851 \pm 0.121	0.829 \pm 0.125	< 0.01
	sd_S1_r1	0.060 \pm 0.030	0.052 \pm 0.028	< 0.01
	m_S1_r2	0.952 \pm 0.080	0.947 \pm 0.072	< 0.01
	sd_S1_r2	0.026 \pm 0.024	0.025 \pm 0.025	0.31
Systolic	m_Sys_r1	0.774 \pm 0.164	0.737 \pm 0.190	< 0.01
	sd_Sys_r1	0.086 \pm 0.039	0.084 \pm 0.047	0.54
	m_Sys_r2	0.892 \pm 0.121	0.859 \pm 0.134	< 0.01
	sd_Sys_r2	0.056 \pm 0.040	0.063 \pm 0.047	< 0.05
S2	m_S2_r1	0.859 \pm 0.118	0.824 \pm 0.121	< 0.01
	sd_S2_r1	0.064 \pm 0.033	0.059 \pm 0.031	< 0.05
	m_S2_r2	0.953 \pm 0.070	0.943 \pm 0.061	< 0.01
	sd_S2_r2	0.029 \pm 0.026	0.030 \pm 0.026	0.15
Diastolic	m_Dia_r1	0.624 \pm 0.192	0.612 \pm 0.220	0.12
	sd_Dia_r1	0.097 \pm 0.041	0.087 \pm 0.037	< 0.01
	m_Dia_r2	0.776 \pm 0.170	0.753 \pm 0.180	< 0.05
	sd_Dia_r1	0.080 \pm 0.041	0.077 \pm 0.042	0.25

Table 2: K=10-fold cross validation results for $\lambda = 1.5$ performed on the balanced dataset: 472 abnormal and 472 normal heart sound recordings. Se: sensitivity, Sp: specificity, Std: standard deviation.

Fold	Training			Validation		
	Se	Sp	Score	Se	Sp	Score
1	0.89	0.89	0.89	0.83	0.60	0.71
2	0.91	0.91	0.91	0.74	0.70	0.72
3	0.90	0.90	0.90	0.72	0.81	0.77
4	0.90	0.90	0.90	0.57	0.74	0.66
5	0.89	0.90	0.89	0.75	0.77	0.76
6	0.92	0.91	0.91	0.81	0.72	0.77
7	0.90	0.89	0.90	0.70	0.68	0.69
8	0.89	0.91	0.90	0.77	0.68	0.72
9	0.90	0.88	0.89	0.70	0.79	0.74
10	0.90	0.92	0.91	0.73	0.75	0.74
Mean	0.90	0.90	0.90	0.73	0.72	0.73
SD	0.01	0.01	0.01	0.07	0.06	0.03