Classifying Heart Sounds
Approaches to the PASCAL Challenge

Elsa Ferreira Gomes[1,4], Peter J. Bentley[2], Miguel Coimbra[3], Emanuel Pereira[4] and Yiqi Deng[2]
[1]*GECAD - Knowledge Engineering Decision Support, Institute of Engineering, Polytechnic of Porto, Porto, Portugal*
[2]*Dept. of Computer Science, UCL, London, U.K.*
[3]*Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Porto, Portugal*
[4]*Institute of Engineering, Polytechnic of Porto, Porto, Portugal*

Keywords: Classifying Heart Sounds, Segmentation, Feature Construction.

Abstract: In this paper we describe a methodology for heart sound classification and results obtained at PASCAL Classifying Heart Sounds Challenge. The results of competing methodologies are shown. The approach has two steps: segmentation and classification of heart sounds. We also describe the data collection procedure.

1 INTRODUCTION

This paper describes the winning approach for the PASCAL Classifying Heart Sounds Challenge. The tasks proposed in this Challenge aim to identify cardiac pathologies by analyzing the features of heartbeat collected from digital stethoscope and from mobile devices. The main components of heart sound signal of a normal heart are the first heart sound, S1 (or lub), corresponding to the systolic period, and the second heart sound, S2 (or dub), the diastolic period (Gupta et al., 2007). This challenge is composed by Challenge 1 (Heart Sound Segmentation) and Challenge 2 (Heart Sound Classification). Attempts to segment phonocardiographic (PCG) signals have been reported in 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 (Groch et al., 1992). Strunic extracted signals on a certain band to reduce anomalies and then set an amplitude threshold to pick out the spikes and perform the segmentation (Strunic et al., 2007). To achieve classification, Karraz extracted the QRS complex from the signal as features and used them in a Neural Network Classifier based on a Bayesian framework (Karraz and Magenes, 2006). Strunic 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 categories. Kampouraki used Support Vector Machines (SVMs) to classify ECG recordings. However, real life data, with varying durations and background noise, is very challenging (Kampouraki et al., 2009). 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 heartbeats, easier (Liang et al., 1997).

2 DATASETS

Dataset A comprises data crowd-sourced from the general public via the iStethoscope Pro iPhone app (Figure 1 - left). iStethoscope Pro is an iOS app which enables members of the public to use their iOS smart phone to listen to their hearts (Palm et al., 2010). The app exploits the excellent audio capabilities of today's mass market devices, performing real-time filtering and amplification, and enabling users to view FFT spectrograms and email 8 seconds of audio. The quality of the audio as assessed by the cardiologists is as good as or better than commercially available digital stethoscopes. Dataset B consists of more than 200 auscultations gathered using the DigiScope Collector system (Figure 1 - right) deployed in the Maternal and Fetal Cardiology Unit of the Real Hospital Português (RHP) in Recife, Brazil (Pereira et al., 2011). Each auscultation consists of 6 to 10 seconds recorded for each of the four standard cardiac auscultation spots in children, which resulted in a to-

F. Gomes E., J. Bentley P., Coimbra M., Pereira E. and Deng Y..
Classifying Heart Sounds - Approaches to the PASCAL Challenge.
DOI: 10.5220/0004234403370340
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2013), pages 337-340
ISBN: 978-989-8565-37-2

tal of 656 audio files provided for the challenge. All relevant patient and auscultation information was annotated by clinicians from RHP using the DigiScope Collector system, including the presence of abnormal sounds such as murmurs. Each individual heartbeat was manually segmented.

Figure 1: iStethoscope and the DigiScope Collector system.

3 CHALLENGE 1 - HEART SOUND SEGMENTATION

In the first challenge we aim to produce a method for determining the location of S1 and S2 sounds within audio data, segmenting the Normal audio existing files in Dataset A and Dataset B. The recorded signals were first preprocessed before performing segmentation. The original signal was decimated, using the decimate function of Matlab (MATLAB, 2010) with factor 5. Then, a band-pass filter was applied. Considering the frequency components of S1 and S2 heart sounds, the chosen filter was a fifth order Chebyshev type I low pass filter with cutoff from 100 Hz to 882 Hz. Then, the signals were normalized to the absolute maximum of the signal (Liang, 1997). After preprocessing, we calculated the Shannon Envelope of the normalized signal. Then, the Average Shannon Energy is calculated in continuous 0.02 seconds windows throughout the signal with 0.01 second overlapping (Liang et al., 1997). After obtaining the normalized average Shannon energy curve we identified the peaks. For that, we adapted the open source function peakdet (Billauer, 2011), written in Matlab. This function finds the local maxima (and minima) using the strategy that a point is considered a maximum peak if it has a locally maximal value, and was preceded (to the left) by a value lower than a given delta. We have used two parameters of the function. The first is the vector to examine, and the second is the peak gap threshold (the delta). In our case, considering this delta on the y values was not enough. We also had to control the distance between peaks on the x-axis, because we know we cannot have two heart sounds too close. Thus, we have changed the function so that a local maximum is considered a peak if the distance to the nearest peak is greater than a second threshold. Using this, we segmented almost all heart sounds. However, we also need to distinguish between S1 and S2, making the correct correspondence to each peak. Our current approach for S1/S2 discrimination is still unsatisfactory. First, we tried to perform the detection of S1 and S2 sounds based on the fact that the distance from S2 to S1 is longer than from S1 to S2, for normal heart rates (Kumar et al., 2006). Bearing this in mind we tried to pick each heart cycle and the corresponding systolic interval. The duration of S1 to S2 segments, or the distance between S1 and S2, was calculated and compared for every segment (Gupta et al., 2007). The longest interval between two sounds was considered to correspond to the diastolic period and the sound at the right side was assigned as S1 and the sound at the left side was assigned as S2. Unfortunately, we find that those intervals vary widely from file to file, in our datasets. This happens because there are very different kinds of heart sound data, for both datasets. In Figure 2 we can see the result of our method for the peak detection. In the 1st chart we have the original signal. In the 2nd chart we have the decimated signal. In the 3rd chart we have decimated signal filtered with a Chebyshev filter. In the 4th chart we have the envelope of the signal and peaks. In the 5th chart we have the peaks over the original signal. For this part of the challenge, the other two teams at the final, the Stanford (Stanford, 2012) and UCL (Deng and Bentley, 2012) teams, used approaches based on wavelet decomposition and spectrogram analysis. To reject the extra peaks, the two teams used the intervals between each adjacent peak (Bentley et al., 2011). The Stanford team uses Shannon energy for the peaks finding. They have used the open source function peakfinder.m, written in Matlab (Yoder, 2009). The results were evaluated on a provided validation set with the correct locations of S1 and S2 sounds. This set contained the segmentation for sounds of the normal category from Dataset A and Dataset B. A test set for final evaluation was also available. This set contained hidden locations but provided the final evaluation results. The total error δ, is calculated by $\delta = \sum_{k=1}^{j} \delta_k$ (Eq. 1).

$$\delta_k = \frac{\sum_{i=1}^{N_k/2}(|RS1_i - TS1_i| + |RS2_i - TS2_i|)}{N_k} \qquad (1)$$

In this equation, δ_k is the average distance of the k-th sound clip in a dataset;N_k is the total number of S1 and S2 in the k-th sound clip; $RS1_i(RS2_i)$ indicates the real location of S1(S2) of the i-th heatbeat and $TS1_i(TS2_i)$ indicates the calculated location of S1(S2) of the i-th heatbeat. j is the total of all the sound clips in the specific dataset. For Dataset B, we obtained a total error of 72242.8 and the other teams obtained 75569.8 (UCL) and 76444.4 (Stanford). For Dataset A the error is higher for all the teams (Bentley et al., 2011). However, in Dataset A Stanford was the best, followed by UCL.

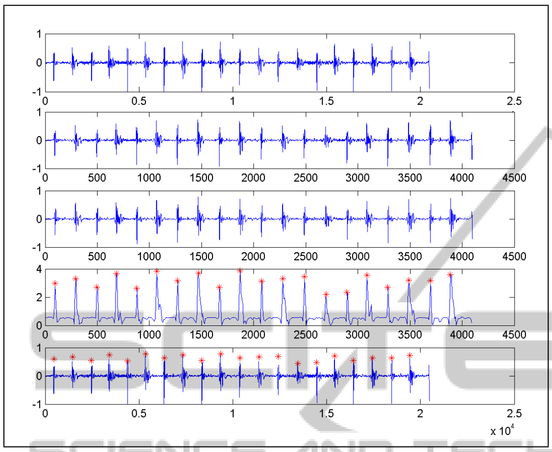

Figure 2: Peak detection on heart sound signal.

4 CHALLENGE 2 - HEART SOUND CLASSIFICATION

The task of Challenge 2 is to produce a method that can classify the real heart audio into one of four classes for Dataset A (Normal, Murmur, Extra Heart Sound and Artifact) and three for Dataset B (Normal, Murmur and Extra systole). This phase involves feature construction and selection and the goal of this phase of the challenge is to label correctly the sounds provided. After the pre-processing and segmentation of the heart sound signal, some features were extracted. Currently, we are using six features. Four of them were extracted from the distances between S1 and S2 (peaks). Assuming that $sS1$ corresponds to smaller segments and $sS2$ to the others, the first feature is the ratio of the standard deviation of $sS1$ over the whole standard deviation. The second is similar for $sS2$. The third and fourth features are the ratio of the mean of $sS1$ ($sS2$ respectively) over total mean. The fifth feature, *Rmedian*, is the ratio of the median of the (three) largest segments in the sample over total mean. The sixth feature, $R2$, is the *r square* of the array of the sorted segments of the sample (a measure of linearity). After obtaining the features we used two different methods from the Weka data mining suite (Witten and Frank, 2005): J48, which generates decision trees, and MLP, the Multi Layer Perceptron.

In Challenge 2, we assess our classification approach using three metrics (per dataset) calculated from the tp (true positives), fp (false positives), tn (true negaties) and fn (false negatives) values. The metrics are precision per class, the Youden's Index, the F-score (only for Dataset A) and the Discriminant Power (only for Dataset B). Precision gives us the positive predictive value (the proportion of samples that belong in category c that are correctly placed in category c). Youden's Index has been used to compare diagnostic abilities of two tests, by evaluating the algorithm's ability to avoid failure. In Dataset A, Youden's Index is evaluated for Artifact category. In Dataset B the Youden's Index is calculated for the problematic heartbeats. In Table 1 and Table 2, we present the results for Dataset A and Dataset B, obtained by our approach after applying the J48 and MLP methods. We also present the results obtained by the UCL team. They focus on the number of heartbeats and features of systole and diastole period, namely on the length of the peak sequence before extra-peak-rejection, the length of finally selected peak sequence after rejection, mean of the systole and diastole period, the standard deviation of diastole period and systolic period. As in Challenge 1, a test set was provided where we could test our method's effectiveness on the unlabelled set.

Table 1: Challenge 2 evaluation for Dataset A.

	J48	MLP	UCL
Precision of Normal	0.25	0.35	0.46
Precision of Murmur	0.47	0.67	0.31
Precision of Extra sound	0.27	0.18	0.11
Precision of Artifact	0.71	0.92	0.58
Artifact Sensitivity	0.63	0.69	0.44
Artifact Specificity	0.39	0.44	0.44
Youden Index of Artifact	0.01	0.13	-0.09
F-Score	0.20	0.20	0.14
Total Precision	1.71	2.12	1.47

Table 2: Challenge 2 evaluation for Dataset B.

	J48	MLP	UCL
Precision of Normal	0.72	0.70	0.77
Precision of Murmur	0.32	0.30	0.37
Precision of Extrasystole	0.33	0.67	0.17
Heart problem Sensivity	0.22	0.19	0.51
Heart problem Specificity	0.82	0.84	0.59
Youden Index of Artifact	0.04	0.02	0.1
Discriminant Power	0.05	0.04	0.09
Total Precision	1.37	1.67	1.31

As we can see in Table 2, our method has problems in classifying the non-normal heartbeats, for Dataset B. In Dataset A, the normal class is one of the most difficult (Table 1). However, we think we can improve our method by improving the correct identification of S1 and S2 in the segmentation and by finding new features that take advantage of this identification.

5 CONCLUSIONS

In this paper, we present the methodology that won the Classifying Heart Sounds PASCAL Challenge. We proposed an algorithm for S1 and S2 heart sound identification (without ECG reference). The segmentation is accomplished by using the envelope of Shannon energy and an algorithm for peak detection. Despite of the good performance for the correct detection of S1 and S2 sounds in the signal, we need to improve the criteria for identifying S1 and S2 (which is which). After the segmentation, we used J48 and MLP algorithms (using Weka) to train and classify the computed features into Normal, Murmur or Extra systole for Dataset B and Normal, Murmur, Extra sound and Artifact for Dataset A. We also compare results obtained by the other two teams present at the final of the competition with ours. Stanford obtained the best results (for Dataset A) on Challenge 1 but did not provide an answer for Challenge 2. Our method, as well as the method followed by the UCL team worked better for Dataset B (the dataset with less noise) than for Dataset A. We think these approaches and this comparative study provide a good basis for further analysis of the heart sound signals. In Challenge 2, our approach with MLP had the highest total precision. Nevertheless, the UCL team performed better in some of the partial success measures.

ACKNOWLEDGEMENTS

We would like to acknowledge the financial support of Fundação para a Ciência e Tecnologia for the DigiScope project with reference PTDC/EIA-CCO/100844/2008. We also thank the PASCAL Network of Excellence for supporting the Classifying Heart Sounds Challenge and Workshop.

REFERENCES

Bentley, P., Nordehn, G., Coimbra, M., and Mannor, S. (2011). The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. www.peterjbentley.com/heartchallenge.

Billauer (2011). www.billauer.co.il/peakdet.html.

Deng, Y. and Bentley, P. J. (2012). A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram. http://www.peterjbentley.com/heartworkshop/challengepaper3.pdf.

Groch, M. W., Domnanovich, J. R., and Erwin, W. D. (1992). A new heart-sounds gating device for medical imaging. *IEEE Transactions on Biomedical Engineering*, 39:307–310.

Gupta, C. N., Palaniappan, R., Swaminathan, S., and Krishnan, S. M. (2007). Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing*, 7:286–297.

Kampouraki, A., Manis, G., and Nikou, C. (2009). Heartbeat Time Series Classification With Support Vector Machines. *IEEE Transactions on Information Technology in Biomedicine*, 13:512–518.

Karraz, G. and Magenes, G. (2006). Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4016–4019.

Kumar, D., Carvalho, R., Antunes, M., Gil, R., Henriques, J., and Eugenio, L. (2006). A New Algorithm for Detection of S1 and S2 Heart Sounds. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 2.

Liang, H., Lukkarinen, S., and Hartimo, I. (1997). Heart sound segmentation algorithm based on heart sound envelogram. *Proceedings of The IEEE*, pages 105–108.

MATLAB (2010). *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.

Palm, D., Burns, S., Pasupathy, T., Deip, E., Blair, B., Flynn, M., Drewek, A., Sjostrand, M., Stephenson, B., and Nordehn, G. (2010). Artificial Neural Network Analysis of Heart Sounds Captured From an Acoustic Stethoscope and Emailed Using iStethoscopePro. *Journal of Medical Devices*, 4(2):027531+.

Pereira, D., Hedayioglu, F., Correia, R., Silva, T., Dutra, I., Almeida, F., Mattos, S., and Coimbra, M. (2011). Digiscope - unobtrusive collection and annotating of auscultations in real hospital environments. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 1193 –1196.

Stanford, S. (2012). http://www.peterjbentley.com/ heartworkshop/challengepaper2.pdf.

Strunic, S. L., Rios-gutirrez, F., Alba-flores, R., Nordehn, G., and Burns, S. (2007). Detection and Classification of Cardiac Murmurs Using Segmentation Techniques and Artificial Neural Networks. In *IEEE Symposium on Computational Intelligence and Data Mining*, pages 128–133.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Yoder, N. C. (2009). www.mathworks.com/matlabcentral.