SIGNAL QUALITY CLASSIFICATION OF MOBILE PHONE-RECORDED PHONOCARDIOGRAM SIGNALS

D.B. Springer[1], T. Brennan[1], L. J. Zühlke[2,3], H. Y. Abdelrahman[2], N. Ntusi[4], G.D. Clifford[1], B.M Mayosi[2], L. Tarassenko[1]

[1]Department of Engineering Science, University of Oxford, United Kingdom
[2]Department of Medicine, Groote Schuur Hospital & University of Cape Town, South Africa
[3]Department of Paediatric Cardiology, Red Cross Children's Hospital & University of Cape Town
[4]Centre for Clinical Magnetic Resonance Research, University of Oxford, United Kingdom

ABSTRACT

There is potential for the use of mobile phones to remotely identify patients with a high risk of heart conditions using automated auscultation. However, accurate heart sound analysis is dependent on the quality of heart sound recordings. This paper investigates the signal quality classification of phonocardiograms (PCGs) recorded on two devices (a 3M Littmann 3200 electronic stethoscope and an iPhone 3G). These recordings were professionally annotated and classified using a support vector machine (SVM) and a combination of ten signal quality metrics computed from each recording as input features. One third of all mobile phone-recorded PCGs were found to be of high quality. The classifier was able to distinguish good and bad-quality iPhone recordings with 87.0% accuracy, the Littmann recordings with accuracy of 76.4% and the combined set with accuracy of 85.6% on unseen test data. Therefore, the quality of PCGs made with a range of stethoscopes can be accurately classified using this technique.

***Index Terms*—** Signal quality, phonocardiogram, mobile health, classification

1. INTRODUCTION

Mobile health has the potential to transform healthcare in the developing world [1] due to the high prevalence of mobile phones [2], the shortage of healthcare workers, poor infrastructure and inadequate training [3], [4].

Rheumatic heart disease (RHD), the leading cause of heart failure in children and young adults worldwide [5], is one condition that could potentially be monitored using a low-cost, mobile stethoscope. This condition, most prevalent amongst children of low socio-economic status in sub-Saharan Africa [6]–[8] results in heart murmurs that are almost always audible during auscultation [9]. A mobile phone-based automatic auscultation device has the potential to identify those individuals with a high risk of having RHD while not requiring expert training or expensive equipment.

This study investigated the feasibility of using a mobile phone to record the heart sounds of patients and classifying the quality of these recordings. Mobile phones could be used by untrained healthcare workers to record heart sounds, or phonocardiograms (PCGs), in the field, provided automated real-time feedback was given to them to ensure the recording of high-quality signals. Automated analysis of such signals could be used to identify high-risk patients in a rural or primary care setting.

2. RELATED WORK

Cardiac auscultation using a mobile phone has previously been investigated. Kuan developed a low-cost home-made stethoscope using a soup ladle (similar to the egg-cup design used in this study) and a low-cost mobile phone hands-free kit [10]. Chen et al. analysed the heart sounds recorded on a HTC G1 and an iPhone 3G [11]. However, these authors did not incorporate any signal quality analysis.

Methods for assessing the signal quality of heart sound recordings have been investigated before. Tosanguan et al. [12] and Kumar et al. [13] both found periods of PCGs that were of low noise. However, they did no classification of whether these recordings were of diagnosable quality.

In only one publication was there the classification of the signal quality of PCG recordings [14]. These authors classified the signal quality of electronic stethoscope recordings by comparing various features against threshold values. However, many important details and values are omitted from their analysis, with the training and testing of their classification performed on the same data.

This study builds on previous work by developing a robust classifier for PCG signal quality, able to differentiate the signal quality of recordings from multiple devices.

3. METHOD

3.1. Data Collection

This study was approved by the Human Research Ethics Committee based at the Health Science Faculty of University of Cape Town (HREC REF: 568/2010). It was conducted in the Cardiac Clinic at Groote Schuur Hospital in Cape Town, South Africa. There were 150 consenting participants recruited for the study and each participant was

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

assigned a unique numeric identifier. Of the 150 patients recruited, 100 had valve replacements (with a mixture of natural and prosthetic valves), 25 had had congestive heart failure, 12 had pacemakers and 11 had congenital disorders.

All recordings were made by a non-medically trained research assistant in order to replicate the role of an untrained healthcare worker. The recordings were performed with the patient in a supine position, after the patient had been resting for five minutes. Sixty second recordings were made with each of the PCG recording devices.

Two different devices were used to record PCG data from each patient: The first was a 3M Littmann 3200 electronic stethoscope. This device recorded PCGs at a sampling frequency of 4 kHz and an amplitude resolution of 16 bits.

The second PCG recording device used was an iPhone 3G mobile phone which recorded PCG data at 44.1 kHz. To ensure adequate coupling between the chest and microphone, a stethoscope attachment, based on the work of Kuan [10], was made by placing the microphone of a standard iPhone hands-free kit into the neck of a metal egg-cup. A rubber seal was attached to the outer rim to ensure an air-tight contact between the cup and the patient's chest. The attachment is shown in Figure 1. The frequency response of the iPhone 3G [15] has an almost flat response even down to low frequencies, leading to minimal distortion, as heart sound frequencies range from 10-1000 Hz [16].

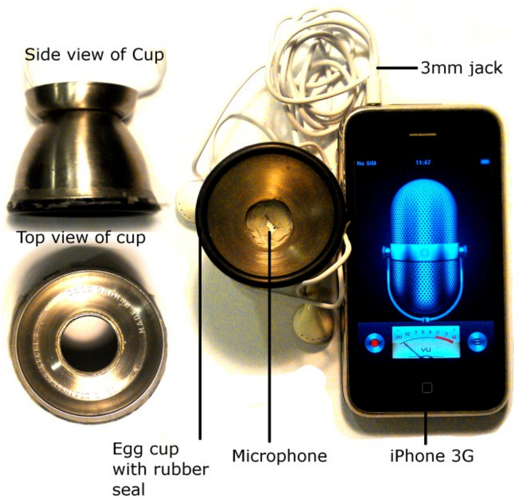

Figure 1: Hand-made stethoscope, made with an aluminium egg-cup and the microphone from a mobile phone hands-free kit plugged into an iPhone 3G.

Table 1: PCG labelling scheme based on diagnostic quality.

Quality Label	Quality Description
1	Excellent - like auscultation. An unequivocal diagnosis can be made, with little to no noise on the recording.
2	Good - like auscultation with noise but still easily heard. Interpretable but some noise present which means expert judgement is needed
3	Borderline - very faint and poorly heard heart sounds and fairly difficult to interpret.
4	Poor - mostly noise with some heart sounds.
5	Awful - no heart sounds, only noise.

3.2. Signal Quality Annotations and Data Exclusion

The PCGs were manually annotated by three cardiologists and one researcher. Each annotator gave each recording a score of one to five based on the diagnostic quality of the recording, using the labelling scheme shown in Table 1. Any PCG with an average annotated score of above two and below four was excluded from further analysis as an ambiguous class. This was done to exclude ambiguous recordings, neither of high or low enough quality to be labelled as such. Recordings with signal quality annotations lower or equal to two were classed as good quality and those higher or equal to four were classed as bad quality for binary classification purposes. The distribution of the annotations for the different recording devices can be seen in Figure 2.

Due to the lack of bad quality Littmann recordings, when these recordings were classified by themselves, recordings with an average annotated score between 1.5 and 2.5 were classed as the ambiguous class, resulting in 66, 57 and 27 good-, ambiguous- and bad-quality recordings respectively. Examples of good and bad-quality recordings on the iPhone can be seen in Figure 3 and Figure 4.

3.3. Preprocessing

Each iPhone PCG recording was down-sampled to 4000 Hz using a polyphase anti-aliasing filter in order to match the Littmann recordings. The frequency content of all heart sounds, normal and pathological, ranges from 10-1000 Hz [16] and hence the Nyquist-Shannon sampling criterion [17] is satisfied.

3.4. Signal Quality Indices

Ten signal quality indices (SQIs), selected from review of the literature, were calculated for each recording. Several of these SQIs were based on the autocorrelation of the envelope of the PCG recording. The envelope of each recording was estimated using the Hilbert transform in a similar manner to [13], [18], [19].

The autocorrelation [20] of the Hilbert envelope was then calculated. The autocorrelation is the cross-correlation of a signal with itself, which accentuates repeating patterns in noisy signals. The autocorrelation waveform was then low-pass filtered using a forward-backward 10 Hz cut-off, second-order infinite impulse response (IIR) low-pass filter, ensuring zero phase distortion. Examples of the autocorrelation waveforms of good- and bad-quality PCG recordings can be seen in Figure 5. The prominent peaks, produced at instances of high correlation between heart sounds, can be clearly identified in the solid line in Figure 5.

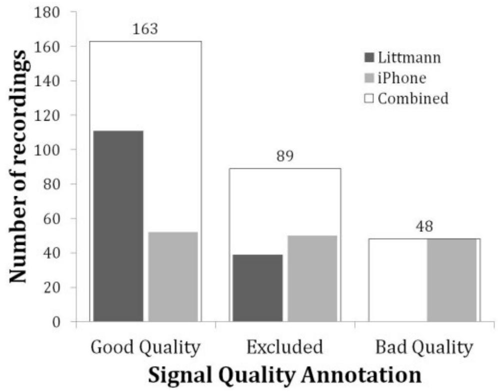

Figure 2: Distribution of the signal quality annotations for the Littmann and iPhone PCG recordings. The total number of good, bad and ambiguous, excluded recordings can be seen.

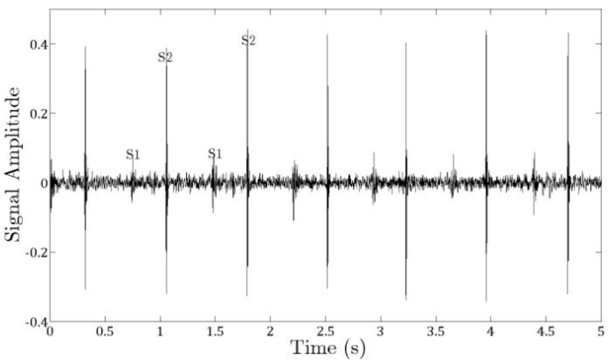

Figure 3: A good-quality iPhone 3G PCG recording with the first (S1) and second (S2) heart sounds identified.

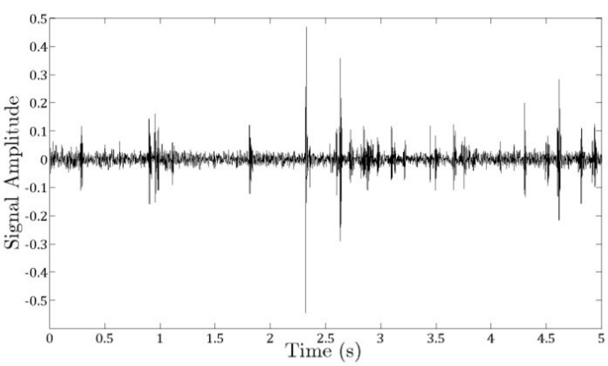

Figure 4: A bad-quality PCG recording made on the iPhone 3G, with large amounts of noise contaminating the signal

The signal quality indices tested for their effectiveness of signal quality classification of the PCG recordings are described in Table 2.

3.5. Support Vector Machine-based Classification

In order to avoid bias in the classification, the two classes (good- and bad-quality) of recordings were randomly split into a training set (2/3 of the recordings) and a test set (1/3 of the recordings), with the same number of good and bad-quality recordings being allocated to each set. In the case of classifying the Littmann and iPhone recordings together, the number of recordings from each device was balanced.

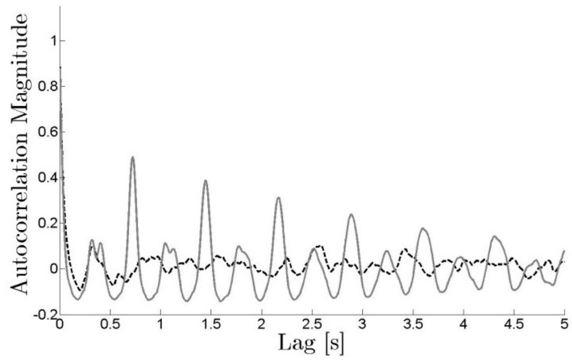

Figure 5: Examples of the autocorrelations of a good- (solid grey line) and bad-quality (dashed black line) PCGs.

Table 2: Description of the ten signal quality indices used in classification of PCG signals

SQI Number	Signal Quality Index	Description
1: m=1, r=0.01 2: m=2, r=0.01 3: m=1, r=0.001 4: m=2, r=0.001	$seSQI_r^m$	The sample entropy [21] of the autocorrelation function, with $m \in \{1,2\}$ being the length of an epoch being measured, while $r \in \{0.01, 0.001\}$ is a threshold value.
5	$kSQI$	The fourth moment (kurtosis) of the autocorrelation waveform (as used in ECG signal quality classification [22]).
6	$svdSQI$	The minimum ratio of the second to first singular value from the singular value decomposition (SVD) of varying window sizes of the autocorrelation function - adapted from [13].
7	$haSQI$	The Hjorth activity [23] or signal power of the autocorrelation function
8	$snrSQI$	The ratio of the signal power in the PCG from 0-240 Hz to 240-1000 Hz.
9	$vSQI$	The variance of the autocorrelation function
10	$ccSQI$	The correlation coefficient between the autocorrelation waveform and a fitted, rectified cosine waveform.

The input features were normalised by subtracting the mean and dividing by the standard deviation for each SQI in the training set. The mean and standard deviation values from the training set were used to normalise the test data.

A support vector machine (SVM) classifier with a Gaussian kernel [24] was used as a classification algorithm, using the libSVM library [25]. The SVM was chosen over other machine learning algorithms as the SVM training procedure always produces a global optimum due to the fact that training the classifier is a convex optimization problem [26].

A SVM with a Gaussian kernel has two parameters, γ, which controls the width of the Gaussian, and C, which controls how strict the classifier is, that can be varied. These were set using cross-validation:

The optimal SVM parameters were found by performing an exhaustive grid search over γ (from 0.1 to 2), C (from 0.5 to 5), the number of features (from 1 to 10) and the combination of features when using leave-two-out (a good- and bad-quality pair of recordings) cross-validation on the training set. The parameters which resulted in the highest average classification accuracy during cross-validation were then used to train the SVM on all the training data. Table 3 shows the number of recordings in the training and test sets for each of the classification groups.

4. RESULTS

The results of the cross-validation optimisation can be seen in Table 4. This table shows the highest average classification accuracy and standard deviation (σ) of the accuracies found over the cross-validation. The standard deviation gives an indication of how widely dispersed the classification results were using the chosen γ, C and features. The average classification results when using these optimised parameters on the separate test set can be seen in Table 5. This table illustrates the classification accuracy on the training set as well as the test set, in order to give an indication of over-training.

5. DISCUSSION AND CONCLUSION

From the distribution of the annotated quality of recordings in Figure 2 it can be seen that 52 of the 150 iPhone recordings were found to be of high, diagnosable quality. This illustrates that a mobile phone equipped with a low-cost stethoscope attachment can be used to record high-quality PCG signals by an untrained healthcare worker.

It can be seen in Table 4 that the classification across datasets is not highly dependent on γ, C, the number and selection of features, as ranges of these resulted in the same classification accuracy. However, the use of between three and eight features across all datasets resulted in the best accuracy. This is expected, due to the risk of over-fitting with small datasets when using many features. The most frequently selected features were sample entropy and the correlation between the autocorrelation and a fitted cosine. These parameters are a measure the periodicity of the repetitive peaks in the autocorrelation function of good quality heart sound recordings.

Table 3: Number of recordings for each classification group

Classification Group	No. of training recordings	No. of test recordings
iPhone 3G	64	32
Littmann	36	18
Combined	64	32

Table 4: Highest average classification accuracy on the leave-two-out cross-validation on the training set, with the features resulting in highest accuracy.

	iPhone	Littmann	Combined
Highest average accuracy (%)	92.2	88.9	96.93
σ of accuracies	18.4	21.4	12.3
γ	1-2	0.6-1.3	1.3-2
C	3-5	2-4	1-3
No. of Features	3-6	3-6	5-8
Most frequently selected features (see Table 2)	1,2,4,6,10	2,4,7,9,10	4,5,6,10

Table 5: Average classification results on train and test sets when using the optimised parameters found on the training sets (%).

		Accuracy	Sensitivity	Specificity
iPhone	Train	96.0	96.8	95.2
	Test	87.0	87.4	86.7
Littmann	Train	93.0	86.2	1
	Test	76.4	63.9	88.9
Combined	Train	97.8	95.7	1
	Test	85.6	87.5	83.4

Table 5 illustrates the success of this technique on the test set. The Littmann recordings had the lowest classification accuracy, explained by the reduced range of good-, ambiguous- and bad-quality data (see Section 3.2) and the difficulty in differentiating between generally good-quality data. The superior training results across all devices indicate the possible presence of over-training.

Therefore, it can be concluded that a mobile phone equipped with a low-cost stethoscope attachment is capable of recording high-quality PCG signals. The poor-quality recordings can be classified with a high degree of accuracy using this technique, which could be used to ensure that an untrained healthcare worker records high-quality PCGs before they are analysed further. However, the classification needs to be made more specific in order to identify all poor-quality recordings, while being careful not to exclude abnormal patient recordings due to poor-quality classification.

A limitation of this work is the reliance of many of the SQIs on the magnitude of peaks in the autocorrelation waveform, which is dependent on the periodicity of the PCG signal. Therefore, any irregular heartbeats, due to arrhythmias or fluctuations in heart rate will affect the SQI values. This could be improved by using beat-to-beat quality metrics, as used in ECG signal quality classification [27].

6. REFERENCES

[1] P. Mechael, “Towards the development of an mHealth strategy. WHO mHealth Review,” 2007.

[2] W. A. Kaplan, “Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries?,” *Global. Health*, vol. 2, no. 9, pp. 1–14, Jan. 2006.

[3] R. M. Scheffler, J. X. Liu, Y. Kinfu, and M. R. Dal Poz, “Forecasting the global shortage of physicians: an economic-and needs-based approach,” *Bull. World Health Organ.*, vol. 86, no. 7, pp. 516–523, 2008.

[4] Y. Kinfu, “The health worker shortage in Africa: are enough physicians and nurses being trained?,” *Bull. World Health Organ.*, vol. 87, no. 3, pp. 225–230, Mar. 2009.

[5] E. Marijon, M. Mirabel, D. S. Celermajer, and X. Jouven, “Rheumatic heart disease.,” *Lancet*, vol. 379, no. 9819, pp. 953–64, Mar. 2012.

[6] E. Marijon, P. Ou, D. S. Celermajer, B. Ferreira, A. O. Mocumbi, D. Jani, C. Paquet, S. Jacob, D. Sidi, and X. Jouven, “Prevalence of rheumatic heart disease detected by echocardiographic screening.,” *N. Engl. J. Med.*, vol. 357, no. 5, pp. 470–6, Aug. 2007.

[7] J. R. Carapetis, A. C. Steer, E. K. Mulholland, and M. Weber, “The global burden of group A streptococcal diseases.,” *Lancet Infect. Dis.*, vol. 5, no. 11, pp. 685–94, Nov. 2005.

[8] B. Longo-Mbenza, M. Bayekula, R. Ngiyulu, V. E. Kintoki, N. F. Bikangi, K. V Seghers, L. E. Lukoki, M. F. Mandundu, M. Manzanza, and Y. Nlandu, “Survey of rheumatic heart disease in school children of Kinshasa town.,” *Int. J. Cardiol.*, vol. 63, no. 3, pp. 287–94, Feb. 1998.

[9] M. K. Park, *The Pediatric Cardiology Handbook*, 3rd ed. St. Louis, MO: Mosby, 2003, p. 311.

[10] K. L. Kuan, “A framework for automated heart and lung sound analysis using a mobile telemedicine platform,” Masters Thesis, Massachusetts Institute of Technology, 2010.

[11] T. Chen, K. Kuan, L. Celi, and G. Clifford, “Intelligent heartsound diagnostics on a cell phone using a hands-free kit,” in *AAAI Spring Symposium on Artificial Intelligence for Development*, 2010, pp. 26–31.

[12] T. Tosanguan, R. J. R. J. Dickinson, and E. M. Drakakis, “Modified spectral subtraction for de-noising heart sounds: Interference Suppression via Spectral Comparison,” in *Biomedical Circuits and Systems Conference*, 2008, pp. 29–32.

[13] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, “Noise detection during heart sound recording using periodicity signatures.,” *Physiol. Meas.*, vol. 32, no. 5, pp. 599–618, May 2011.

[14] H. Naseri and M. R. Homaeinezhad, “Computerized quality assessment of phonocardiogram signal measurement-acquisition parameters.,” *J. Med. Eng. Technol.*, vol. 36, no. 6, pp. 308–18, Aug. 2012.

[15] GSM Arena Team, “Apple iPhone 3GS review: Same clothes, new feel,” 2009. [Online]. Available: http://www.gsmarena.com/apple_iphone_3gs-review-369p6.php#aq. [Accessed: 01-Nov-2013].

[16] A. K. Abbas and R. Bassam, *Phonocardiography Signal Processing*. Morgan & Claypool Publishers, 2009, p. 194.

[17] C. E. Shannon, “Communication In The Presence Of Noise,” *Proc. IEEE*, vol. 86, no. 2, pp. 447–457, Feb. 1998.

[18] S. R. Messer, J. Agzarian, and D. Abbott, “Optimal wavelet denoising for phonocardiograms,” *Microelectronics J.*, vol. 32, no. 12, pp. 931–941, Dec. 2001.

[19] A. Sarkady, R. Clark, and R. Williams, “Computer analysis techniques for phonocardiogram diagnosis,” *Comput. Biomed. Res.*, vol. 9, pp. 349–363, 1976.

[20] G. Box, G. M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting & Control (3rd Edition)*. Prentice Hall, 1994, p. 592.

[21] T. A. Kuusela, T. T. Jartti, K. U. O. Tahvanainen, and T. J. Kaila, “Nonlinear methods of biosignal analysis in assessing terbutaline-induced heart rate and blood pressure changes.,” *Am. J. Hear. Circ. Physiol.*, vol. 282, no. 2, pp. H773–83, Feb. 2002.

[22] Q. Li, R. G. Mark, and G. D. Clifford, “Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter.,” *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, Jan. 2008.

[23] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Amsterdam: Elsevier, 2005, p. 688.

[24] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[25] C. Chang and C. Lin, “LIBSVM: a library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006, p. 740.

[27] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, “Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms.,” *Physiol. Meas.*, vol. 33, no. 9, pp. 1419–33, Sep. 2012.