# Preliminary Results on Quantification of Seismocardiogram Morphological Changes, Using Principal Component Analysis

Vahid Zakeri, Kouhyar Tavakolian, Siamak Arzanpour, John M. Zanetti, Guy A Dumont, and Alireza Akhbardeh

***Abstract*—A methodology, based on principal component analysis, is proposed to quantify beat to beat Seismocardiogram changes. The proposed method was tested over a population of 94 subjects including 35 ischemic heart disease patients. The results showed that there was an insignificant overlap between the diseased and the healthy populations in the number of principal components (NPC) and that further development of this method might yield a classification index for myocardial abnormalities. In addition such an index has potential utility in patient monitoring.**

## I. Introduction

Every heartbeat sets the body into mechanical vibrations. These vibrations have been recorded using different methodologies for the past century and have been given different names based on the recording site or measuring technology. Two very important types of such mechanical signals are Ballistocardiograms (BCG) and Seismocardiograms (SCG). BCGs are essentially created by the movement of the center of gravity of a body through blood circulation, whereas SCGs are created by local vibrations of the chest, and are recorded using accelerometers [1]. We consider Seismocardiogram the low frequency component of the accelerograms recorded from the chest. If the same accelerograms are high-pass filtered (25 Hz) they correspond to the phonocardiogram [2]. Fig. 1 compares 5 SCG segments from a healthy and a diseased subject. As Fig. 1 indicates, there is greater similarity between SCG segments of the healthy subject.

There have been some attempts to quantify such similarities. Issac Starr, a pioneer of BCG research, came up with a qualitative BCG classification from beat to beat morphology changes and then used this classification to separate subjects into four classes. In the first class, all BCG complexes are normal in contour. In the 2nd class, the majority of the complexes are normal, but one or two of the smaller complexes of each respiratory cycle are abnormal in contour. In the 3rd class the majority of the complexes are abnormal in contour, usually only a few of the largest complexes of each respiratory cycle remaining normal and in the 4th class there is such complete distortion that the waves cannot be identified with confidence and the onset of ejection could not be located without a simultaneous ECG recording [3].

Some preliminary results on quantification of beat to beat morphological changes of BCG was reported using signal to noise ratios [4]. The same group took a similar approach for quantification of morphological changes in processing the higher frequency component of accelerograms recorded from the chest. These components correspond to phonocardiogram's $S_1$ and $S_2$ complexes [5].

They assumed that each complex could be modeled as an underlying "template" function that was scaled in amplitude from one segment to another. Their model included a time-delay representing the interval between $S_1$ and $S_2$, as well as considering the unknown measurement noise. They estimated the signals with their proposed model, and indicated that the obtained signal-to-noise ratio (SNR) of healthy subjects was more than diseased subjects. Using such an approach is based on hypothesizing a fixed template for different components of $S_1$ and $S_2$ complexes. Such a hypothesis may not always be satisfied. Moreover, the validation of their approach was conducted on a small number of subjects.

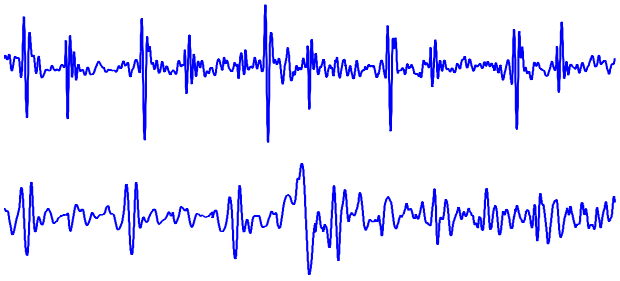

Figure 1. Five SCG segments of a healthy subject (top plot) and an unhealthy one diagnosed with ischemic heart disease (bottom plot). The similarity between SCG segments of the healthy subject is more.

In this paper, a method for quantifying the similarity among SCG segments is introduced that is not dependent on morphological conditions. This method is introduced in section II, and is tested on a database of SCG signals consisting of 59 healthy subjects, and 35 with ischemic heart disease.

Resrach was supported by the engage grant from Natural Sciences and Engineering Research Council of Canada (NSERC).

Vahid Zakeri is with the School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, BC, Canada, e-mail: vahid_zakeri@sfu.ca.

Kouhyar Tavakolian is with the Department of Electrical Engineering, University of North Dakota, Grand Forks, ND, USA, e-mail: kouhyar@und.edu.

Siamak Arzanpour is with the School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, BC, Canada, e-mail: arzanpour@sfu.ca.

John M. Zanetti is with Acceleron Medical Systems, Wisconsin, USA. email: jmzsenior@gmail.com

Guy A Dumont is with Department of Electrical Engineering, University of British Columbia, Vancouver, BC, Canada, email: guyd@ece.ubc.ca

Alireza Akhbardeh is with the School of Medicine, Johns Hopkins University, Baltimore, MD, USA, email: akhbardeh@jhu.edu

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

## II. METHODS

In order to quantify the similarity between SCG segments, our proposed method included different stages as in Fig 2. These stages are discussed in the following sections.

### A. Subjects and Data Acquisition

The SCG signal was measured with a high sensitivity piezoelectric accelerometer (Brüel & Kjær model 4381, Nærum, Denmark) using a National Instrument data acquisition system with sampling frequency of 1000 Hz. The participants were in the supine position and the signals were recorded in back-to-front direction, perpendicular to the body surface. The Electrocardiogram (ECG) signal was simultaneously recorded.

The patient population (n=35, Age=66.1±8.2 years, Weight=84.2±21.19 kg, Height=169.8±8.6 cm) were recruited from healthy heart program at Burnaby General Hospital under an ethics approval from Fraser Health Authorities. These patients were diagnosed with ischemic heart disease (IHD) and were being monitored by a cardiologist at the hospital to improve their cardiac performance.

The healthy population (n=59, Age=32.1±4.7, Weight=79±11, Height=173.4 ± 7.2) were selected mainly from Simon Fraser University students and staff under an ethics approval from the university and informed consent was documented for every subject.

### B. Segmentation and Preprocessing

The SCG signals were first segmented using their corresponding ECG signals. Each segment of SCG was considered from R to R peaks of ECG, but shifted by 200 samples to the left. After segmentation, the preprocessing was conducted on each signal. In this stage, each segment was normalized to have a zero-mean and a unit variance.

### C. Length Equalization and Alignment

For each SCG signal, the lengths of all segments were equalized to their maximum length using zero padding. The SCG segments then aligned using Woody's approach [6]. Woody's is an iterative algorithm that aligns the segments based on the maximum correlation between each segment and the average of segments. The average can be updated in each iteration to reach the desired alignment accuracy.

### D. Quantifying the Similarity

In order to quantify the similarity between SCG segments, principal component analysis (PCA) was chosen. PCA is a very popular method for analysing the variations among different data points. PCA finds the direction that maximizes the variation of projected data [7].

To employ PCA, the data are first organized in a matrix $D$, then the covariance matrix $C$ is computed,

$$C = \frac{1}{m-1}(DD^T) \tag{1}$$

where $D^T$ denotes transpose of matrix D, and m is the length of each segment. Finally, the eigenvalues ($\lambda$) and eigenvectors ($\upsilon$) of D are computed as follow,

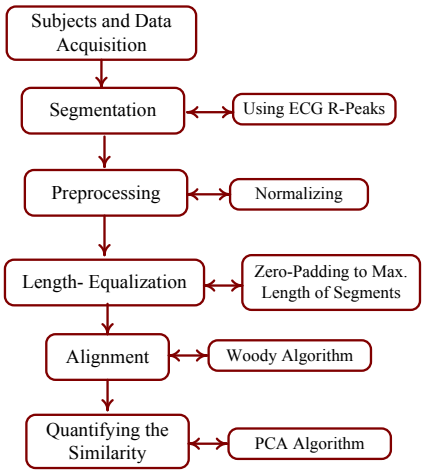


Figure 2. The schematic diagram of the proposed method for quantifying the similarity between SCG segments

$$C\upsilon_i = \lambda_i \upsilon_i \qquad i = 1, 2, \dots, m \tag{2}$$

In the Eq. (2), it is assumed that eigenvalues are sorted in the descending order, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Considering this assumption, the principal components P (in the descending order) can be computed as,

$$P_i = D\upsilon_i \qquad i = 1, 2, \dots, m \tag{3}$$

The principal components $P_1$, are the projections of the original data in the direction that have the maximum variance. $P_2$ have the maximum variance after $P_1$, and so forth for $P_3, \dots, P_m$. The set of all principal components is as large as the original set of data. However, it is common to choose the number of principal components whose sum of variances reaches a portion ($\alpha$) of the total variance of the original data. The number of principal components (NPC) can be computed as,

$$\sum_{i=1}^{NPC} \mathrm{Var}(P_i) = \alpha \sum \mathrm{diag}(C) \qquad 0 < \alpha < 1 \tag{4}$$

In the Eq. (4), $\mathrm{diag}(C)$ represents the diagonal elements of the covariance matrix C, which are the variances of the original data.

Basically, NPC is a variable that can quantify the similarity between SCG segments. A large value for NPC indicates more variation (more dissimilarity); whereas a small value shows fewer variations (more similarity).

For this study, the SCG signals of 94 subjects were investigated; in which 59 were healthy, and 35 were diseased. The segments of each SCG signal were organized in a matrix $D_{m\times n}$, where m and n indicated the length of each segment, and the number of segments respectively. PCA algorithm was employed using MATLAB platform. The number of principal components (NPC) was selected for $\alpha = 80\%$ according to Eq. (4). NPC was computed for all the subjects, and Fig. 3 shows a histogram of the number of subjects versus NPC. Fig. 4 displays a box-plot of NPC indicating max, min, median, percentiles, outliers, and range of data both for healthy and diseased subjects.

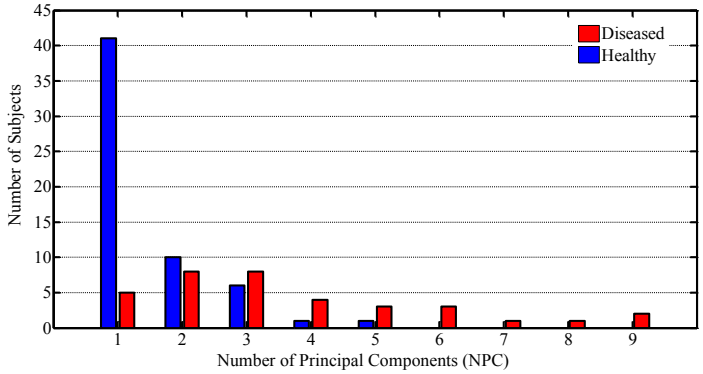


Figure 3. The histogram of number of subjects versus NPC. Smaller values of NPC indicate more similarity between segments. Most of healthy segmenst have NPC = 1, and higer values of NPC has occurred for diseased subjects.

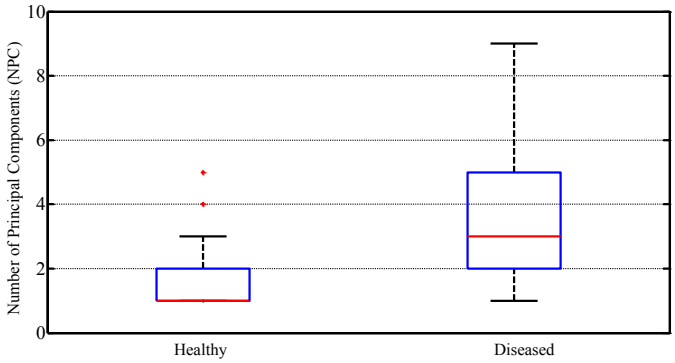


Figure 4. The boxplot of healthy and diseased subjects indicating, max, min, median, percentiles, outliers, and range of data.

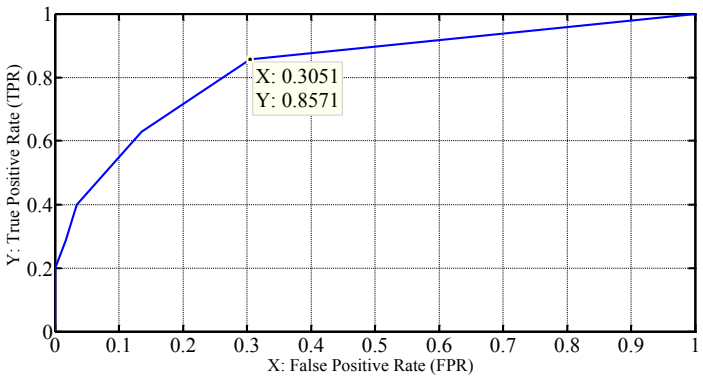


Figure 5. The receiver operation characteristic (ROC) curve for classification between healthy and diseased subjects

As Fig. 3 shows, most of the healthy subjects (41 out of 59) had NPC = 1, and higher values of NPC such as 6 – 9 only belonged to diseased subjects. For other subjects, by increasing the NPC value, the number of healthy subjects reduced whereas the number of diseased subjects increased (e.g. NPC = 3 and in comparison with NPC = 1, and NPC = 9). Using boxplot, Fig. 4 shows the range of NPC value for diseased subjects to be more compared to healthy ones. In addition, the maximum NPC for diseased subjects (9) was approximately two times greater than the maximum (5) for healthy ones.

Another useful analysis is using the receiver operation characteristic (ROC) curve [8] which is depicted in Fig. 5. This curve shows the true-positive-rate (TPR) versus the false-positive-rate (FPR) for a binary classifier with different thresholds THR,

$$\mathrm{THR} = 1, 2, \ldots, 9 \tag{5}$$

In each threshold THR, the binary classifier assigns a positive class (+1) for NPC values greater than or equal to THR. In this analysis, it is assumed that diseased subjects are in the positive class (+1), and healthy subjects are in the negative class (-1). The ideal point in the ROC plane is (FPR=0, TPR=1) which corresponds to 100% classification accuracy for both classes. As Fig. 5 indicates, the nearest point on ROC curve to the ideal point was obtained for (FPR= 0.3051, TPR= 0.8571). This point corresponded to THR=2, in which the classification accuracies were 85.7% and 69.5% for diseased and healthy subjects respectively.

## III. Discussion and Conclusion

In this paper, a method for quantifying the similarity between SCG segments was introduced. In our method, the number of principal components (NPC) was proposed as a quantity that

could indicate the similarity. Basically, a lower value of NPC showed less variance or more similarity between SCG segments, whereas, a higher value of NPC indicated more dissimilarity. Our proposed method is applicable to any signal with a periodic structure; however in this study we employed it on SCG signals to differentiate between healthy and diseased subjects. It should be noted that there was an age difference among the healthy and diseased subjects, which also could contribute to morphological changes. The effect of such an age difference on the value of NPC should be independently investigated in future studies by age matching the two populations.

Our hypothesis was that the similarity between segments of healthy subjects was greater than the diseased ones. Our results confirmed this hypothesis, and NPC indicated a very good separation among healthy/diseased subjects.

## ACKNOWLEDGEMENT

We would like to thank Burnaby General Hospital for providing access to the patients and Heart Force Medical Inc. for providing the data acquisition system.

## REFERENCES

[1] K. Tavakolian, B. Ngai, A. P. Blaber, and B. Kaminska, "Infrasonic Cardiac Signals : Complementary," in *IEEE EMBC*, 2011, vol. 6, pp. 4275–4278.

[2] J. M. Zanetti and K. Tavakolian, "Seismocardiography: Past, present and future," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2013, pp. 7004–7007.

[3] A. Starr, I, Noordergraaf, *Ballistocardiography in Cardiovascular Research*. Philadelphia: Lippincott Company, 1967.

[4] L. Giovangrandi, O. T. Inan, D. Banerjee, and G. T. A. Kovacs, "Preliminary results from BCG and ECG measurements in the heart failure clinic," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, vol. 2012, pp. 3780–3783.

[5] O. T. Inan, K. Pandia, L. Giovangrandi, R. T. Zamanian, and G. T. A. Kovacs, "A preliminary study investigating the quantification of beat-to-beat in seismocardiogram signals," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, vol. 2013, pp. 7286–7289.

[6] C. D. Woody, "Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals," *Med. Biol. Eng.*, vol. 5, no. 6, pp. 539–554, Nov. 1967.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[8] M. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.