



**HAL**  
open science

## Sensor Data Quality Processing for Vital Signs with Opportunistic Ambient Sensing

Ibrahim Sadek, Jit Biswas, Zhu Yongwei, Zhang Haihong, Jayachandran Maniyeri, Chen Zhihao, Teo Ju Teng, Ng Soon Huat, Mounir Mohktari

► **To cite this version:**

Ibrahim Sadek, Jit Biswas, Zhu Yongwei, Zhang Haihong, Jayachandran Maniyeri, et al.. Sensor Data Quality Processing for Vital Signs with Opportunistic Ambient Sensing. The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society , Aug 2016, Orlando, FL, United States. 10.1109/EMBC.2016.7591234 . hal-01336329

**HAL Id: hal-01336329**

**<https://hal.science/hal-01336329>**

Submitted on 23 Jun 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sensor Data Quality Processing for Vital Signs with Opportunistic Ambient Sensing

Ibrahim Sadek, *Member, IEEE*, Jit Biswas, Zhu Yongwei, Zhang Haihong, Jayachandran Maniyeri, Chen Zhihao, Teo Ju Teng, Ng Soon Huat, and Mounir Mokhtari

**Abstract**—Opportunistic ambient sensing involves placement of sensors appropriately so that intermittent contact can be made unobtrusively for gathering physiological signals for vital signs. In this paper, we discuss the results of our quality processing system used to extract heart rate from ballistocardiogram signals obtained from a micro-bending fiber optic sensor pressure mat. Visual inspection is used to label data into informative and non-informative classes based on their heart rate information. Five classifiers are employed for the classification process, i.e., random forest, support vector machine, multilayer, feedforward neural network, linear discriminant analysis, and decision tree. To compute the overall effectiveness of quality processing, the informative signals are processed to estimate interbeat intervals. The system was used to process, data collected from 50 human subjects sitting in a massage chair while performing different activities. Opportunistically collected data was obtained from the fiber optic sensor mat placed on the headrest of the massage chair. Using our classification approach, 57.37% of the dataset was able to provide informative signals. On the informative signals, random forest classifier achieves the best classification accuracy with a mean accuracy of 98.99%. The average of the mean absolute error between the estimated heart rate and the reference ECG is reduced from 13.2 to 8.47. Therefore, the proposed system shows a good robustness for opportunistic ambient sensing.

## I. INTRODUCTION

Opportunistic Ambient Sensing (OAS) may be used to provide applications and services that fit into active and healthy lifestyle of end users, and to unobtrusively extract reliable and meaningful data about their physiological parameters [1]. Although mobile phone based applications are perhaps the most convenient, they lack the sensitivity and the proximity needed for the measurement of important vital signs such as heart rate (HR) and breathing rate (BR). Ambient sensors that are placed in the environment away from physical contact with the user, such as cameras, infrared motion sensors and other types of electromagnetic sensing devices, lack the proximity or physical contact needed for reliable detection of vital signs.

The micro-bending fiber optic sensor (FOS), thanks to its high sensitivity to ballistic effects of human vital signs, is

Ibrahim Sadek is with Image and Pervasive Access Lab (IPAL), CNRS UMI 2955, Singapore, Télécom ParisTech, Institut Mines Télécom, France, Institute for Infocomm Research (I<sup>2</sup>R), Singapore (email: stuis@i2r-astar.edu.sg)

Jit Biswas, Zhu Yongwei, Zhang Haihong, Jayachandran Maniyeri, Chen Zhihao, Teo Ju Teng, Ng Soon Huat are with Institute for Infocomm Research (I<sup>2</sup>R), Singapore (email: {biswas, ywzhu, hhzhang, mjay, zchen, jtteo, shng}@i2r.a-star.edu.sg)

Mounir Mokhtari is with Image and Pervasive Access Lab (IPAL), CNRS UMI 2955, Singapore, Institut Mines Télécom, France (email: mounir.mokhtari@mines-telecom.fr)

a sensor that is suitable for opportunistic ambient sensing. This sensor is a proper choice for non-intrusive continuous monitoring because it is highly responsive to pressure changes caused by body movement, and simultaneously not required to be in close contact with the body. A cushion with embedded sensors, for example, is able to unobtrusively, yet with great accuracy, capture vital signs of the user for those durations where the user is directly in contact with the sensor and motion artifacts are limited.

A major challenge is dealing with the quality of the FOS signals. There are two basic approaches to monitoring quality in physiological data obtained from sensors, i.e., signal oriented and aggregate oriented. The first [2] is an exact approach, attempting to detect and track statistical properties of signal morphologies or event occurrences (or non-occurrences), and reporting these in a real-time fashion. The second [3], [4] is a statistical approach aimed at obtaining aggregate statistical features in the time or frequency domain, through appropriate feature extraction. Since the former approach can lead to processing delay, especially in real-time situations, in our work we adopt the statistical approach for quality processing.

The goal of this research paper is to design a quality processing system to identify signals of interest from the noisy and non-stationary ballistocardiogram (BCG) signals. The quality process reduces the computational and / or communications load significantly, because only useful data is transmitted and processed for vital sign extraction. Thereafter, interbeat intervals are estimated using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) algorithm.

The paper is summarized as follows. Section II presents the generic algorithm for quality processing. Section III presents a definition of data quality relevant to the needs of the application at hand. Section IV briefly presents the data collection protocol, a labeling tool used to manually label datasets accordingly, and the proposed approach and Section V presents the results of the quality processing system. Conclusions are presented in Section VI.

## II. QUALITY PROCESSING OF BCG SIGNALS FROM FIBER OPTIC SENSOR PRESSURE MAT

Ballistocardiography is a technique to measure the vibrations of the body during the rapid ejection of blood from the left ventricle into the large blood vessels synchronous with each heartbeat [5]. Fig. 1 shows the FOS used for the purpose of our study. It includes a pressure mat embedded with

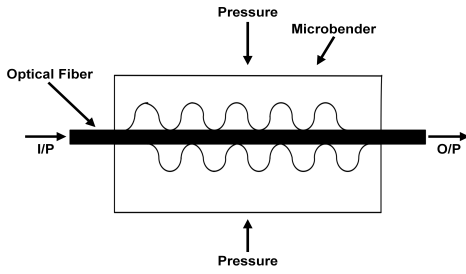


Fig. 1. Micro-bending fiber optic sensor.

micro-bending sensors and a transceiver, which is connected to a computer via Bluetooth. When a force is exerted on the mat, the short distance between a pair of micro-benders alters. Consequently, the light intensity in the multi-mode fiber changes according to the ballistic forces on the heart. This force modulates the light intensity and then retrieved as a ballistocardiogram signal. The micro-bending FOS pressure

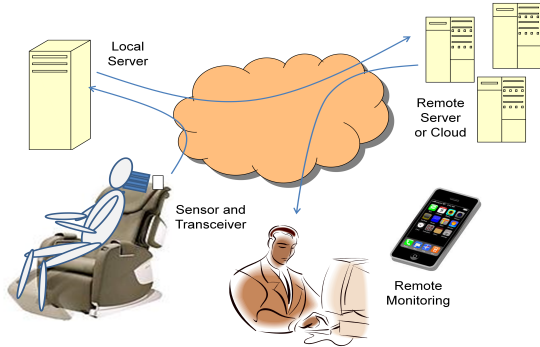


Fig. 2. Opportunistic remote monitoring of vital signs.

mat [6], [7], [8], [9] can be used to gather physiological data of people from ambient locations, i.e., mattresses, pillows, chairs, etc. In these situations, a main obstacle is to handle noise and motion artifacts, while the quality of the sensed data is easily corrupted. Recovering vital signs with high accuracy can be achieved if some features of the signals can be utilized. The algorithm necessitates carrying out labeling of sensor data, in order to establish reliable ground truth, which can then be used to segment good quality signals from real-time sensor data streams. This is a step towards improving the reliability of sensors used to monitor vital signs in diverse ambient settings. Fig. 2 illustrates the manner in which the FOS pressure mat is deployed in wellness applications for the opportunistic monitoring of vital signs.

### III. APPLICATION SPECIFIC SIGNAL DATA QUALITY

Sensor signal quality and data quality are important measures for medical devices and health monitoring. In morphological analysis, a missed detection is characterized by the absence of a particular waveform. The missing or erroneous waveform is corrected and the feature (e.g., heartbeat) is then detected from the augmented waveform. In statistical analysis, the individual waveforms are not analyzed, but the signal is processed and feature detected (e.g., through peak detection) and thereafter the output statistic (e.g., HR) computed. The statistical approach at data quality requires

manual labeling of the rendering of the signal on a viewer with the help of a suitable tool. The waveforms are classified, sometimes with the help of logged information that supply details that may be useful for more accurate labeling. Since the fiber optic sensor can be used in various settings for various applications, the statistical approach towards the study of quality is more meaningful, as it may be used with parametric or algorithmic changes and applied to various applications and indeed to various sensors as well. For the remainder of this paper, we will consider the following definitions of quality:

- **Informative signal:** It includes a good quality signal and a noisy signal. The first is a data stream from which features may be extracted by standard algorithms, without any further filtering or processing to remove noise or any conflicting artifacts. The second is a data stream that contains the observed signal mixed with other artifacts and noise, such that more processing needs to be done before algorithms may be used to extract the vital signs.
- **Non-informative signal:** It is a data stream where retrieval of physiological properties is impossible.

Because noise is introduced through activities of various kinds, in the experimental stage we incorporate mechanisms to embed information into the data that indicate the occurrence of events. These labels give major cues, which serve as a foothold in the data, in order to analyze it further for quality processing. Note that even though our data is collected in the lab, it is unreasonable and impossible to label each event. This is the reason why we need to undertake post processing of the data through a manual labeling effort. From Fig. 2, it can be seen that there are certain periods when good data is expected (such as the rest periods), and certain other periods where the data is expected to be noisy (such as the massage period when the massage chair is in motion). There are also periods when, due to non-contact there is no signal presented. This happens, for example, when the subject is answering the questionnaire or undergoing some mental tasks involving the use of the computer. Once the subject leans forward the contact between the pillow sensor and the subject's head is lost, and thus there is no signal.

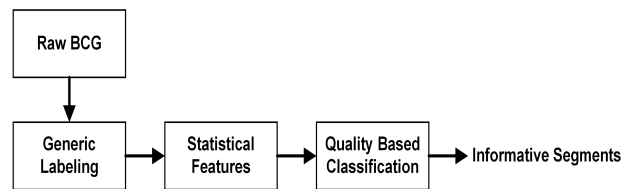


Fig. 3. Flowchart of the quality processing system.

## IV. METHODOLOGY

### A. Data Collection

The data is collected in a realistic setting from 50 human subjects sitting in a massage chair to assess their levels of stress at different moments. Consent forms were sought from every human subject following the approval of the

institutional review board. Participants completed stress-inducing exercises, proceeded by rest and relief therapy. During the process, subjects were instructed to complete well-suited survey questionnaires for gathering self-reported ground truth. Meanwhile, physiological parameters of subjects were recorded in real-time using a range of sensors. This includes FOS, EEG, galvanic skin response, and some other sensors to measure ECG and respiratory efforts.

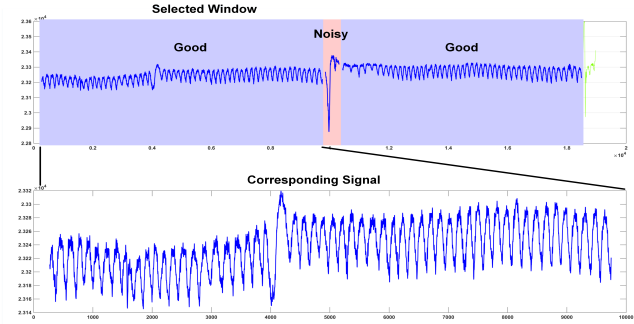


Fig. 4. Screen capture of the labeling tool.

### B. Classification

The flowchart of the quality processing system is shown in Fig. 3 and can be illustrated as follows:

- 1) Preprocessing: The sensor data is manually labeled by a human observer into two classes such as 1) informative (58% of dataset) and 2) non-informative (42% of dataset), where a labeling tool was developed that enables the user to quickly label data from MATLAB<sup>®</sup> based software. Fig. 4 shows a screen capture of the labeling tool. The labeled sensor data is randomly divided into two groups, i.e., G1 and G2. The former consist of 2085 segments (1296 informative and 789 non-informative), whereas the latter includes 1546 segments (813 informative and 733 non-informative). The length of each segment is 10 seconds/500 samples, where the data is sampled at a sampling frequency of 50Hz. The idea is to use G1 as a training set while G2 as a test set, and vice versa. Subsequently, each segment is band pass filtered to extract the BCG component using a Butterworth band-pass filter with frequency limits of 1Hz and 12Hz. Fig. 5 shows two examples for informative and non-informative segments.
- 2) Feature Extraction: A set of 13 statistical features is extracted i.e., mean, standard deviation, minimum, maximum, skewness, kurtosis, range, interquartile range, median absolute deviation, number of zero crossings, variance of local maxima, variance of local minima, mean of the signal envelope.
- 3) Training and Testing: Five classifiers are employed, i.e., random forest (RF), support vector machine (SVM), multilayer, feedforward neural network (NN), linear discriminant analysis (LDA), and decision tree (DT). Then, a training model is created for each classifier using the features of training set, where each

set is 10-fold cross validated to evaluate the predictive ability of the models. Finally, each segment in the test set is classified into one of the classes based on the features of the test set. Furthermore, an accuracy criterion is computed for appraising the performance of the classifiers and the results of the proposed approach in discriminating between informative and non-informative segments.

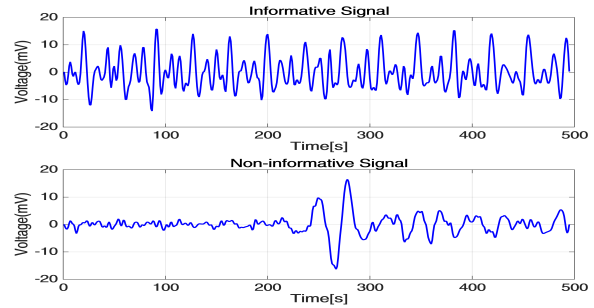


Fig. 5. Two examples for informative and non-informative segments.

- 4) Heart Rate Estimation: The heart beats are detected as presented in [10], since the CEEMDAN is proved to be robust towards motion artifacts induced because of body movements, while reducing the computational complexity. The 6<sup>th</sup> decomposition component is chosen for HR measurement because each local maximum shows a match for cardiac cycle. Fig. 6 shows an example of a BCG signal with its 6<sup>th</sup> decomposition component.

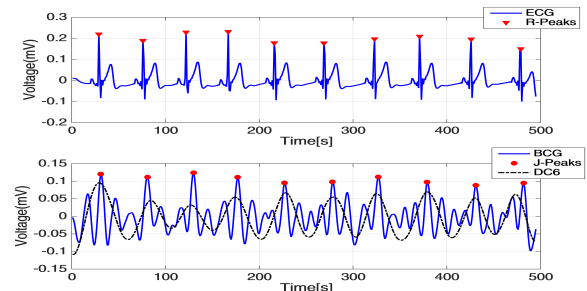


Fig. 6. BCG signal with a reference ECG signal. ECG is shown in 1<sup>st</sup> row. However, BCG signal and its 6<sup>th</sup> decomposition component are shown in 2<sup>nd</sup> row.

## V. RESULTS AND DISCUSSION

As mentioned in the previous section, the goal is to use a cross dataset testing in which G1 is to be used as a training set while G2 as a test set and vice versa. Therefore, the selected classifier should be able to correctly classify the data of G2 based on G1 and contrariwise. Following the 10-fold cross validation of each group, the best parameters of the classifiers are identified and the mean accuracy of each classifier is determined. Table I shows the mean accuracy of the 10-fold cross validation for each classifier. In both groups, random forest classifier shows superior performance over the other classifiers with a mean accuracy of 98.13% and 92.30% for G1 and G2 respectively. Using the same classifier, similar

TABLE I

10-FOLD CROSS VALIDATION MEAN ACCURACY FOR G1 AND G2, (RF: NTREES = 50, SVM: RBF KERNEL, NN: 50 HIDDEN NEURON).

		Mean Accuracy (%)				
		RF	SVM	NN	LDA	DT
G1		98.13	93.38	91.61	89.26	97.51
G2		92.30	90.49	85.89	79.37	89.39

TABLE II

ACCURACY RESULTS FOR TESTING G2 Vs. G1 (EXP1) AND TESTING G1 Vs. G2 (EXP2).

		Accuracy (%)				
		RF	SVM	NN	LDA	DT
Exp1		100	94.44	92.28	89.40	99.42
Exp2		97.99	97.46	87.10	90.26	97.41
Average		98.995	95.95	89.69	89.83	98.41

results are achieved when G2 is tested versus G1 and the other way around with an accuracy of 100% and 97.99% for Exp1 and Exp2 accordingly as presented in Table II. Table III shows the confusion matrix of random forest for Exp2. Moreover, it can be included that the percentage of recovered informative data is 57.37%, which can be calculated as follows. Percentage =  $(1270 + 813)/(2085 + 1546)$ . Decision tree and support vector machine classifiers attain quite similar results to random forest with accuracy results of (99.42% & 97.41%) and (94.44% & 97.46%) for Exp1 and Exp2 respectively.

TABLE III

CONFUSION MATRIX OF RANDOM FOREST FOR TESTING G1 Vs. G2 (EXP2).

		Actual	
		Informative	Non-Informative
Predicted	Informative	1270	26
	Non-Informative	18	771

On the other hand, the CEEMDAN is applied to BCG signals following a noise standard deviation of 0.35, an ensemble size of 100, and a maximum number of iterations of 30 to obtain interbeat intervals. For ECG and BCG segments, the HR is estimated in beat per minute and the mean absolute error (MAE) is used to assess the performance of the CEEMDAN algorithm. In order to estimate the effectiveness of the quality processing system, HR is estimated before and after applying the classification process. From Table IV, it

TABLE IV

THE MAE FOR G1 AND G2 BEFORE AND AFTER CLASSIFICATION.

	MAE (before)	MAE (after)
G1	11.16	7.26
G2	15.24	9.68
Average	13.2	8.47

can be noted that, the MAE is largely reduced from (G1: 11.16 and G2: 15.24) to (G1: 7.26 and G2: 9.68) after classification. As a result, the quality processing system can

effectively increase robustness of the system for vital signs monitoring.

## VI. CONCLUSION AND FUTURE DIRECTIONS

In wellness applications, the necessity for sensor data processing is not immediate. However, opportunistic sensing can be used. The activity considered in this paper is resting on a massage chair, with the sensor embedded into the headrest of the massage chair. The algorithm presented makes use of machine learning techniques to process the quality of sensor data, thereby filtering out only significant segments of the signal. Random forest classifier achieves the best classification performance to identify informative and non-informative signals with respect to their heart rate information. The heart rate is detected using CEEMDAN algorithm, where each local maximum of the 6<sup>th</sup> decomposition component corresponds to the periodicity of the J-Peaks of the BCG signal. The performance of the CEEMDAN is determined according to MAE measurement. The proposed quality processing system plays a key role to eliminate the erroneous of the system to monitoring vital signs. In the future, we would like to propose a more robust approach to compute heart beats in addition to respiratory efforts.

## REFERENCES

- [1] G. Acampora, D. Cook, P. Rashidi, and A. Vasilakos, "A survey on ambient intelligence in healthcare," *Proceedings of the IEEE*, vol. 101, pp. 2470–2494, Dec 2013.
- [2] J. Schumm, *Quality assessment of physiological signals during ambulatory measurements*. Ph.D. dissertation, ETH Zurich, 2010.
- [3] J. Wang, "A new method for evaluating ecg signal quality for multi-lead arrhythmia analysis," in *Computers in Cardiology, 2002*, pp. 85–88, Sept 2002.
- [4] S. Lee, C. Ling, A. Nahapetian, and M. Sarrafzadeh, "A mechanism for data quality estimation of on-body cardiac sensor networks," in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pp. 194–198, Jan 2012.
- [5] O. Inan, P.-F. Migeotte, K.-S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G. Prisk, and M. Di Rienzo, "Ballistocardiography and seismocardiography: A review of recent advances," *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, pp. 1414–1427, July 2015.
- [6] Z. Chen, X. Yang, J. T. Teo, and S. H. Ng, "Noninvasive monitoring of blood pressure using optical ballistocardiography and photoplethysmograph approaches," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 2425–2428, July 2013.
- [7] Y. Zhu, H. Zhang, M. Jayachandran, A. Ng, J. Biswas, and Z. Chen, "Ballistocardiography with fiber optic sensor in headrest position: A feasibility study and a new processing algorithm," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 5203–5206, July 2013.
- [8] J. Biswas, Y. Zhu, H. Zhang, J. Maniyeri, Z. Chen, and C. Guan, "Information processing of optical sensor data in ambient applications," in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on*, pp. 1–6, April 2014.
- [9] Z. Chen, J. T. Teo, S. H. Ng, X. Yang, B. Zhou, Y. Zhang, H. P. Loo, H. Zhang, and M. Thong, "Monitoring respiration and cardiac activity during sleep using microbend fiber sensor: A clinical study and new algorithm," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pp. 5377–5380, Aug 2014.
- [10] I. Sadek, J. Biswas, V. F. S. Fook, and M. Mokhtari, "Automatic heart rate detection from fbg sensors using sensor fusion and enhanced empirical mode decomposition," in *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*, pp. 349–353, Dec 2015.