AWARD NUMBER: W81XWH-12-1-0541

TITLE: Wearable Wireless Sensor for Multi-Scale Physiological Monitoring

PRINCIPAL INVESTIGATOR: Ki H. Chon

CONTRACTING ORGANIZATION: Worcester Polytechnic Institute
Worcester, MA 01609

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2014	Annual	25 Sep 2013 - 24 Sep 2014

4. TITLE AND SUBTITLE

Wearable Wireless Sensor for Multi-Scale Physiological Monitoring

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0541

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Ki Chon
Yitzhak Mendelson

E-Mail: kichon@wpi.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

One of the aims of Year 2 of the project was to complete development of a prototype multi-channel pulse oximeter that can be used to collect physiological data from multiple body locations to combat motion artifact contamination. Specifically, the aim was to investigate if a motion artifact-free signal can be obtained in at least one of the multi-channels at any given time. Towards this aim, we have developed a prototype 6-photodetector reflectance-based pulse oximeter and results to date show that good signals can be obtained in one of the multi-channels at any given time. These devices are currently in use for field testing in our labs and at UMASS. Moreover, it was found that both forehead- and ear-located pulse oximeters provide better signal quality than a finger pulse oximeter. The second major aim of the project was to develop a motion and noise detection algorithm and a separate algorithm for the reconstruction of the motion and noise contaminated portion of the data. For detection of motion and noise artifacts, we have successfully developed an accurate and real-time realizable algorithm. For all new data including that from UMASS, our new MNA algorithms consistently provide better accuracy than our previously-published algorithm. A manuscript describing this new algorithm has been submitted for publication. In the past year, we have published 4 journal articles and 2 journal articles have been submitted based on algorithm development. Moreover, we have filed 2 new patent disclosures on our algorithms. Finally, we have recruited 105 patients to date at UMASS. These data will be used in the year 3 to investigate the robustness of our motion and noise artifact algorithms. Both the sensor and algorithms will be thoroughly tested and further refined, if needed, using the UMASS data collected in Year 3 of the project.

15. SUBJECT TERMS
Motion and noise artifact detection and reconstruction; multi-channel pulse oximeter sensor

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	82	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

US combat experience has demonstrated that acute hemorrhage and subsequent hemodynamic decompensation (*shock*) account for about 50% of the deaths on the battlefield. Realizing the limits of current pre-symptomatic diagnosis and treatment capabilities on the battlefield, a reliable non-invasive physiological sensor and diagnostic algorithms that provide *clinical decision support* for early hemorrhage diagnosis and facilitate *remote assessment (triage)* for medical evacuation of the highest-priority combat casualties remains one of the primary objectives for Combat Casualty Care. Moreover, a sensor that can monitor the status of uninjured soldiers suffering from physiologic stress such as dehydration, may help optimize performance in the field. To address this challenging deficiency and reduce the medical logistics burden in the field, we propose to significantly enhance the current capabilities of our prototype wearable, pulse oximeter-based, physiological status sensor so that when donned by military personnel it will acquire and wirelessly transmit in real-time ***seven*** algorithmically derived vital physiological indicators: heart rate, perfusion index, oxygen saturation, respiratory rate, autonomic nervous system dynamics, arrhythmia detection, and blood volume loss. This critical information will be captured, analyzed and displayed on a hand-held monitoring device carried by a medic. Any change in a soldier's physiological status including early warnings of impending hemorrhagic shock or severe dehydration will alert the individual responsible for monitoring soldiers' conditions so that appropriate timely intervention may be taken. Our sensor will be applicable in at least two different scenarios: remote combat triage and bedside (point of care) monitoring. For the latter scenario, our recently developed smart phone technology which uses images processed from a fingertip to derive seven physiological parameters using our algorithms is also applicable. Our ***single*** sensor (either wearable pulse oximeter itself or pulse oximeter-like information derived from a smart phone) combines significant advancements in both sensors and patent pending detection algorithms that are especially applicable for accurate and early detection of hemorrhage on spontaneously breathing subjects, a feat that has not been achieved to date.

Keywords

Motion and noise artifacts, pulse oximeter, photoplethysmogram, smart phone, hypovolemia, vital sign, hemorrhage, wearable sensors, time-varying, time-frequency, support vector machine

Overall Project Summary

Our results from both withdrawing 900 ml blood and a lower body negative pressure study to simulate significant blood loss suggest the potential use of the photoplethysmogram signal for early diagnosis and quantification of hypovolemia at levels of blood loss earlier than can be identified by changes in vital signs or physician estimation. This is a particularly novel and highly relevant use of the PPG in detection of blood loss, considering the fact that vital signs may not show discernable changes even up to 30% blood volume loss. We have four **Specific Aims**:

1) To develop miniaturized multi-channel pulse oximeter hardware that can be used from multiple body locations (forehead, chest and wrist) to combat motion-artifact contamination.
2) To develop motion artifact detection and removal software utilizing photoplethysmogram signals acquired by the multi-channel pulse oximeter which will provide clean signals for the calculation of seven physiological parameters, including hypovolemia. In addition, we will develop and test an application for derivation of the seven parameters from a pulsatile signal gathered with an Android-based smart phone camera.
3) To evaluate the ability of our physiologic sensor to detect acute blood volume loss and monitor resuscitation in a prospectively recruited group of trauma patients presenting to the emergency department.
4) To evaluate the ability of our physiologic sensor to detect significant intravascular and total body volume depletion/dehydration and monitor resuscitation in a prospectively recruited cohort of patients presenting to the emergency department.

Key Research Accomplishments

Aim1

- We developed prototype multi-channel forehead, finger and ear pulse oximeter sensors.
- These devices are currently in use for field testing in our labs and at UMASS.
- It was found that both forehead- and ear-located pulse oximeters provide better signal quality than a finger pulse oximeter.
- We have demonstrated that better quality PPG signals can be obtained from a multi-channel pulse oximeter when compared to a single channel pulse oximeter sensor.

Aim 2

- We developed a new MNA detection algorithm that is more accurate and computationally efficient than our previously-developed method. For all new data including that from UMASS, our new MNA algorithms consistently provide better accuracy than our previously-published algorithm. A manuscript describing this new algorithm has been submitted for publication.
- We developed 2 new algorithms for reconstructing MNA-corrupted portions of the data. These manuscripts are in preparation and will be submitted in the 9th quarter. These 2 new algorithms are more computationally efficient than our recently-published algorithm.

- We have converted one of the 2 new reconstruction algorithms into C language and real-time implementation is now possible on a personal computer. We have demonstrated this capability to the Program Managers during the MHSRS meeting this past August.
- These new algorithms are able to reconstruct not only time-varying changes in heart rates but they also capture the PPG amplitudes with good precision.
- We have implemented these algorithms in a smart phone application. While the algorithms are more computationally efficient than other published algorithms, we had to simplify the algorithms to make them functional in real-time. However, this is not a good option as we had to reduce some functionality, thus, we will continue to optimize these algorithms in the current year to make them viable for real-time application without compromising the accuracy.
- We have published 4 journal articles (3 articles in Annals of BME rated as one of the top journals in BME with an impact factor of 3.2, and 1 article in the journal Sensors, impact factor of 2.02) and 2 journal articles have been submitted based on algorithm development.
- We have filed 2 new patent disclosures on our algorithms.
- We developed a new method to obtain tidal volume from a smartphone without using any external sensors; a patent disclosure has been filed on this new technology.

Aims 3&4

- Our initial patient enrollment plan was the following: 20 control, 30 dehydration and 30 trauma subjects. However, we increased these targets to: 80 control, 60 dehydration and 80 trauma subjects. This increase in subject enrollments was approved by both UMASS and HRPO.
- To date we have 105 patients enrolled in our study.
- Data analysis on these subjects has just begun.

Conclusions

In Year 2 of our project, we have completed development of prototype multi-channel pulse oximeters that can be used to collect physiological data from multiple body locations (forehead, finger and ear lobe) to combat motion-artifact contamination. These multi-channel pulse oximeters were used to collect clinical data in Year 2 of the project at the Emergency Department of the University of Massachusetts Medical Center as well as at WPI. Concurrent with the sensor development, we have successfully developed more robust algorithms for the detection followed by reconstruction of the identified motion and noise corrupted data segments. From these efforts, we have published 4 journal articles, submitted 2 additional articles, and 3 other articles are currently in preparation to submit to journals. Moreover, we have submitted 2 new patent disclosures to WPI on new algorithms for robust and accurate detection and reconstruction of motion and noise artifact contaminated data. Further, a new patent disclosure was filed for estimation of tidal volume using only the video images from a smartphone. In Year 3, we will continue to investigate the robustness of our motion and noise artifact algorithms on UMASS-collected data. Both the sensor and algorithms will be thoroughly tested and further refined, if needed, using the UMASS data collected in Year 3 of the project. We have also received verbal confirmation from Dr. Convertino that he will be sending us some field data which will certainly contain rich motion and noise artifacts. We are currently on schedule for all milestone tasks originally proposed in our grant proposal.

Publications, Abstracts, and Presentations

Refereed Journal Articles

1. Chong, J, D.K. Dao, S.M.A. Salehizadeh, D.D. McManus, C.E. Darling, **K.H. Chon**, and Y. Mendelson, Photoplethysmograph signal reconstruction based on a novel hybrid motion artifact dection-reduction approach- Part I: motion and noise artifact detection, *Annals of BME*, 42:2238-2250, 2014.
2. Salehizadeh, S.M.A., D.K. Dao, J. Chong, Y. Mendelson, D.D. McManus, C. Darling and **K.H. Chon**, Photoplethysmograph signal reconstruction based on a novel motion artifact detection- reduction approach- Part II: motion and noise artifact removal. *Annals of BME*, 42:2238-2250, 2014.
3. Reyes, B., N. Reljin, and **K.H. Chon**, Tracheal sounds acquisition using smartphones, *Sensors, 14*(8): 13830-13850, 2014.
4. Nam, Y., J. Lee and **K.H. Chon**, Respiratory rate estimation from the built-in camera of smartphones or tablets. *Annals of BME*, 42:885-898, 2014.
5. Lazaro, J., Y. Nam, E. Gil, P. Laguna, and **K.H. Chon**, Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals, Under Review.
6. Dao, D.K., J.W. Chong, S.M.A. Salehizadeh, C.H. Cho, D.D. McManus, C. Darling, Y. Mendelson and **K.H. Chon**, A robust motion artifact detection algorithm for photoplethysmogram signals using time-frequency spectral features, Under Review.

Conference Proceedings:

1. Lazaro, J., Y. Nam, E. Gil, P. Laguna and K.H. Chon, Smartphone-camera-acquired pulse photoplethysmographic signal deriving respiratory rate, 8th Conference of the European Study Group on Cardiovascular Oscillations, ESGCO, 2014, Trento, Italy.

Inventions, Patents and Licenses

1. Estimation of Tidal Volume Using Video Camera Images of Smartphones, Invention disclosure filed with WPI. Inventors: Ki Chon, Y. Nam and Bersain Reyes
2.
3. Motion and Noise Artifact Detection and Reconstruction Algorithms for Photoplethysmogram Signals, Invention disclosure filed with WPI. Inventors: Ki Chon, Duy Dao, Hamed Salehizadeh and Jowoon Chong.

Reportable Outcomes

- 4 journal articles published as noted above and 2 journal articles under review.
- 2 new patent disclosures
- Developed prototype multi-channel PPG sensors for forehead, finger and ear.
- Developed real-time motion and noise artifact detection and reconstruction algorithms for demonstrations using a person computer.
- Developed a smartphone application for real-time detection of motion and noise artifacts
- Developed a smartphone application for real-time data collection of video camera images from a smartphone for tidal volume estimation.

Other Achievements

None

References

None

Appendix

4 published journal articles are provided.

Annals of Biomedical Engineering (© 2014)
DOI: 10.1007/s10439-014-1080-y

BMES BIOMEDICAL ENGINEERING SOCIETY™ www.bmes.org

Photoplethysmograph Signal Reconstruction Based on a Novel Hybrid Motion Artifact Detection–Reduction Approach. Part I: Motion and Noise Artifact Detection

JO WOON CHONG,[1] DUY K. DAO,[1] S. M. A. SALEHIZADEH,[1] DAVID D. MCMANUS,[2] CHAD E. DARLING,[3] KI H. CHON,[1] and YITZHAK MENDELSON[1]

[1]Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA; [2]Cardiology Division, Departments of Medicine and Quantitative Health Sciences, University of Massachusetts Medical Center, Worcester, MA 01655, USA; and [3]Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA 01655, USA

(*Received 28 February 2014; accepted 25 July 2014*)

Associate Editor Tingrui Pan oversaw the review of this article.

Abstract—Motion and noise artifacts (MNA) are a serious obstacle in utilizing photoplethysmogram (PPG) signals for real-time monitoring of vital signs. We present a MNA detection method which can provide a clean vs. corrupted decision on each successive PPG segment. For motion artifact detection, we compute four time-domain parameters: (1) standard deviation of peak-to-peak intervals (2) standard deviation of peak-to-peak amplitudes (3) standard deviation of systolic and diastolic interval ratios, and (4) mean standard deviation of pulse shape. We have adopted a support vector machine (SVM) which takes these parameters from clean and corrupted PPG signals and builds a decision boundary to classify them. We apply several distinct features of the PPG data to enhance classification performance. The algorithm we developed was verified on PPG data segments recorded by simulation, laboratory-controlled and walking/stair-climbing experiments, respectively, and we compared several well-established MNA detection methods to our proposed algorithm. All compared detection algorithms were evaluated in terms of motion artifact detection accuracy, heart rate (HR) error, and oxygen saturation (SpO_2) error. For laboratory controlled finger, forehead recorded PPG data and daily-activity movement data, our proposed algorithm gives 94.4, 93.4, and 93.7% accuracies, respectively. Significant reductions in HR and SpO_2 errors (2.3 bpm and 2.7%) were noted when the artifacts that were identified by SVM-MNA were removed from the original signal than without (17.3 bpm and 5.4%). The accuracy and error values of our proposed method were significantly higher and lower, respectively, than all other detection methods. Another advantage of our method is its ability to provide highly accurate onset and offset detection times of MNAs. This capability is important for an automated approach to signal reconstruction of only those data points that need to be reconstructed, which is the subject of the companion paper to this article. Finally, our MNA detection algorithm is real-time realizable as the computational speed on the 7-s PPG data segment was found to be only 7 ms with a Matlab code.

Keywords—Motion and noise artifacts, Photoplethysmography, Support vector machine.

INTRODUCTION

PPG is a non-invasive and low cost device to continuously monitor blood volume changes in peripheral tissues.[28] PPG is a useful technique since it is widely used to monitor HR, SpO_2, and can also be used to measure respiratory rates.[24] However, MNA can distort PPG recordings, causing erroneous estimation of HR and SpO_2.[28,32] It is because of the MNA that PPG has not yet been widely adopted as a possible sensor for mobile health applications. There are three distinct sources of MNA that can distort PPG recordings: environmental, physiological and experimental artifacts, which can be attributed to (1) electromagnetic and power interference around the body; (2) cross talk pickup of other physiological signals; and (3) instrumental noise, respectively.[18,36,39] MNA, which are comprised of all of the aforementioned noise sources, are difficult to filter since they do not have a predetermined frequency band and their spectrum often overlaps with that of the desired PPG signal. Despite

Address correspondence to Ki H. Chon, Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA. Electronic mail: jw9607@gmail.com, dkdao2013@gmail.com, smasalehizadeh@gmail.com, mcmanusd@ummhc.org, Chad.Darling@umassmed.edu, kichon@wpi.edu, ym@wpi.edu

Published online: 05 August 2014

these challenging scenarios, in our companion paper, we describe a method that can reconstruct the MNA contaminated data segments so that accurate heart rates and SpO_2 values can be estimated.

MNA in PPG readings are caused by (1) the movement of venous blood as well as other non-pulsatile components along with pulsatile arterial blood and (2) variations in the optical coupling between the sensor and the skin.[2,28,37,38] Various approaches to mitigate motion artifacts by improving sensor attachment have been proposed.[20,27] However, these design improvements do not provide a significant reduction of motion artifacts. Algorithm-based MNA reduction methods were also proposed. These include time and frequency domain filtering, power spectrum analysis, and blind source separation techniques.[12,15,19,29–31,40] However, these have high computational complexity and more importantly, they operate even on clean PPG portions where MNA reduction is not needed and consequently may distort the clean signal.[15–17,29] Hence, accurate MNA detection, which identifies clean PPG recordings from corrupted portions, is essential for the subsequent MNA reduction algorithm so that it does not distort the non-corrupted data segments.[16] Moreover, more computationally efficient MNA algorithms can be designed since they can be tailored only to the MNA contaminated data segments.

MNA detection methods are mostly based on a signal quality index (SQI) which quantifies the severity of the artifacts. Some approaches quantify SQI using waveform morphology[21,22,35] or filtered output,[14] while other derive SQI with the help of additional hardware such as accelerometer and electrocardiogram sensing.[7,17] Statistical measures, such as skewness, kurtosis, Shannon entropy, and Renyi's entropy, have been shown to be helpful in determining a SQI.[5,41] However, these techniques require manual threshold settings for each parameter to classify if the PPG signal is clean or corrupted. Although a support vector machine (SVM)-based classification method addresses the need of threshold setting,[41] this approach considers limited and controlled types of motions. The authors are not aware of any detailed studies providing representative and comprehensive features distinguishing clean from corrupted PPG signals under various types of motions.

In this paper, we propose an accurate and comprehensive MNA detection algorithm which detects MNA in PPG under various types of motion. We first introduce time-domain parameters quantifying MNA in the recorded PPG signal. We then consider their statistical measures as input variables for a machine learning-based MNA detection algorithm. Our MNA detection algorithm is self-trained by the SVM with clean and corrupted PPG data sets, and then the trained SVM tests the unknown PPG data. We tested the efficacy of our proposed algorithm on PPG data sets recorded from the finger and forehead pulse oximeters in simulations, laboratory-controlled and walking/stair-climbing experiments, respectively.

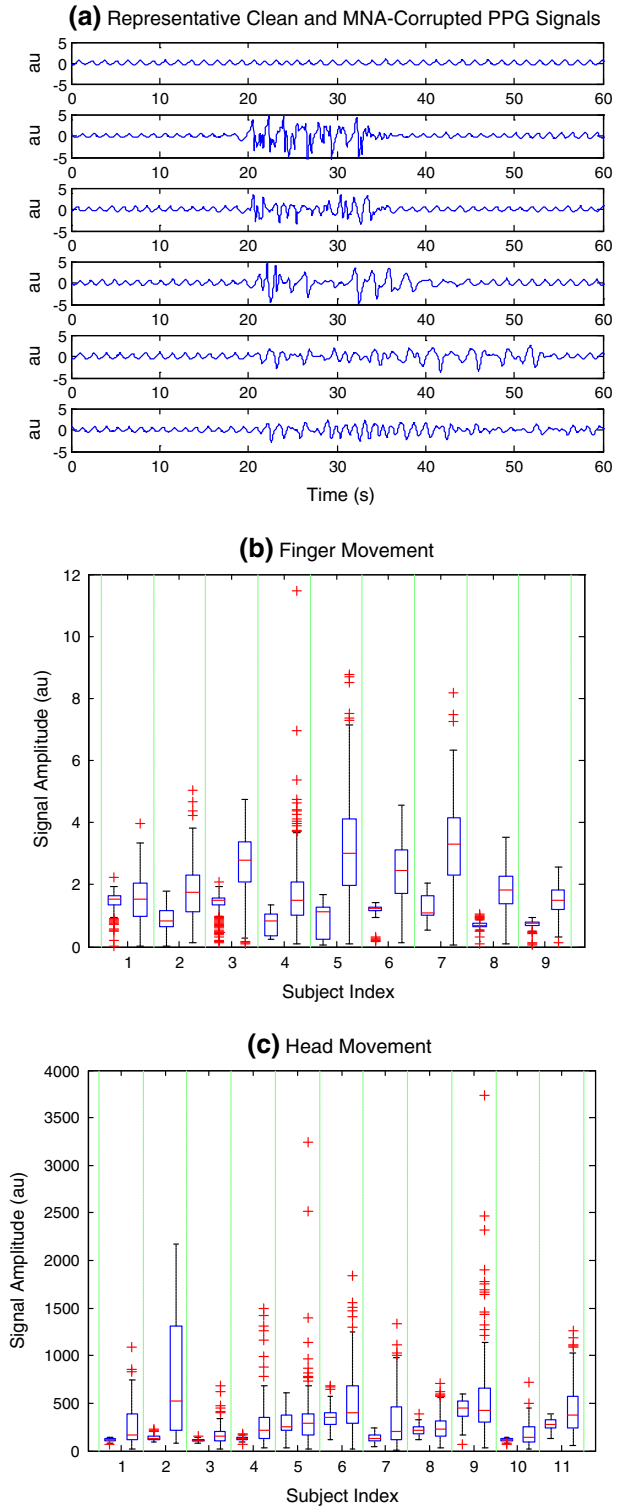
MATERIALS AND METHOD

Experimental Protocol and Preprocessing

PPG signals were obtained from custom reflectance-mode prototype pulse oximeters. PPG data with laboratory-controlled head and finger movement, daily-activity movement, or simulated movement were collected respectively from healthy subjects recruited from the student community of Worcester Polytechnic Institute (WPI). This study was approved by WPI's IRB and all subjects were given informed consent prior to data recording.

In *laboratory-controlled head movement* data, motion artifacts were induced by head movements for specific time intervals in both horizontal and vertical directions. Eleven healthy volunteers were asked to wear a forehead reflectance pulse oximeter along with a reference Masimo Radical (Masimo SET®, Masimo Corporation, CA, USA) finger type transmittance pulse oximeter. After baseline recording for 5 min without any movement, subjects were instructed to introduce motion artifacts for specific time intervals varying from 10 to 50% within a 1 min segment as shown in Fig. 1a. For example, if a subject was instructed to perform left–right movements for 6 s, a 1 min segment of data would contain 10% noise. MNA amplitudes varied for each subject as shown in Fig. 1b. Specifically, mean amplitude ratios and the deviation ratios (between 75th mean and 25th percentile values) between corrupted and clean signals are in the ranges of [0.9935 3.0092] and [1.3375 11.1250], respectively. The right middle finger with the sensor attached to the Masimo pulse oximeter was kept stationary. HR and SpO_2 signals were acquired by the Masimo pulse oximeter at 80 and 1 Hz, respectively, and were acquired synchronously with the PPG signals recorded from the forehead sensor.

In *laboratory-controlled finger movement* data, motion artifacts were induced by left–right movements of the index finger. Nine healthy volunteers were asked to sit and wear two reflection type PPG pulse oximeters (TSD200, BIOPAC Systems Inc., CA, USA) on their index and middle fingers, respectively. After baseline recording for 5 min without any movement to acquire clean data, motion artifacts were induced by left–right movements of the index finger while the middle finger was kept stationary as a reference. Similar to the head



◀FIGURE 1. PPG signals recorded during voluntary motion artifact conducted in a laboratory setting. (a) A representative clean forehead-PPG signal recorded during voluntary motion artifact conducted in a laboratory setting (1st row). The mixed (up–down and left–right) movement of the forehead to which the PPG probe was attached for predetermined time interval induced 10–50% noise (2nd–6th row) within a 60 s PPG segment (b) amplitudes of clean (left) and MNA-corrupted (right) finger-PPG signals of 9 subjects, and (c) amplitudes of clean (left) and MNA-corrupted (right) forehead-PPG signals of 11 subjects. The central line on each box corresponds to the median; the edges of the box correspond to the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, and outliers are plotted individually.

movement data, motion was induced at specific time intervals corresponding to 10–50% duration in an 1 min segment. Such controlled movement was repeated five times per subject. MNA amplitudes varied for each subject as shown in Fig. 1c. The mean amplitude ratios and deviation ratios (between 75th mean and 25th percentile values) between corrupted and clean signals are in the range of [0.9360 4.0226] and [1.3892 32.2647], respectively. The pulse oximeters were connected to a biopotential amplifier (PPG100, BIOPAC Systems Inc., CA, USA) having a gain of 100 and cut-off frequencies of 0.05–10 Hz. The MP1000 (BIOPAC Systems Inc., CA, USA) was used to acquire finger PPG signals at 100 Hz. The *daily-activity movement PPG data* were recorded while subjects were walking straight or climbing stairs for 45 min. Nine subjects were asked to walk or climb stairs after wearing a forehead reflectance pulse oximeter along with a Holter electrocardiogram (ECG) monitor (Rozinn RZ153+, Rozinn Electronics, Inc., NY, USA) at 180 Hz and a Masimo Rad-57 pulse oximeter (Masimo Corporation, CA, USA) at 0.5 Hz. The reference ECG was obtained from the Holter ECG monitor while HR and SpO_2 readings were measured from the Masimo pulse oximeter connected to the subject's right index finger, which was held against the chest to minimize motion artifacts. Finally, the *simulation movement PPG data* were generated by the addition of white noise to the clean PPG data. PPG data were preprocessed by a 6th order infinite impulse response (IIR) band pass filter with cut-off frequencies of 0.5 and 12 Hz. Zero-phase forward and reverse filtering was applied to account for the non-linear phase of the IIR filter. After these preprocessing, the following parameters for classifying clean and corruption were derived.

Parameters from PPG Signals

Motion-corrupted PPG signals are observed to have noticeably different pulse amplitudes, pulse widths and trough depths when compared to clean PPG signals.[35] Moreover, morphology and amplitude ratios of corrupted PPG signals differ from those of clean signals. It was found that most PPG signals show strong similarity among noise-free pulses, but large variation among successive poor and bad quality pulses.[35] Based on our observations on measured clean and corrupted

PPG signals shown in Fig. 1 as well as previous work,[35] we considered standard deviation values of the pulse amplitude, pulse width and pulse shape to quantify differences among these features. We also considered systolic and diastolic ratios since they are observed to be within a well-defined range for clean PPG pulses.[23] The following four parameters were selected since they represent the variability present in corrupted PPG signals as shown in Fig. 1.

Standard Deviation of Peak-to-Peak Interval (STD_{HR})

The $STD_{HR,n}$ of the nth segment is defined by:

$$STD_{HR,n} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(D_{n,i} - \overline{D_n}\right)^2} \quad (1)$$

where $D_{n,i}$ is peak-to-peak interval at the ith pulse of the nth segment and $\overline{D_n}$ is mean peak-to-peak interval of the nth segment. The $D_{n,i}$ is calculated by the difference $T_{peak,n,i} - T_{peak,n,i-1}$ between two successive peak times.

Standard Deviation of Peak-to-Peak Amplitude (STD_{AMP})

The $STD_{AMP,n}$ is calculated by substituting $D_{n,i}$ and $\overline{D_n}$ in (Eq. 1) with $A_{n,i}$ and $\overline{A_n}$, respectively, where $A_{n,i}$ is peak amplitude at the ith pulse of the nth segment and $\overline{A_n}$ is mean peak-to-peak interval of the nth segment. The $A_{n,i}$ is defined by the difference between the ith peak and the forthcoming $(i+1)^{th}$ trough amplitude values.

Standard Deviation of Systolic and Diastolic Ratio (STD_{SD})

The $STD_{SD,n}$ of the nth segment is derived by replacing $D_{n,i}$ and $\overline{D_n}$ in (Eq. 1) by $R_{SD,n,i}$ and $\overline{R_{SD,n}}$, respectively, where $R_{SD,n,i}$ is systolic and diastolic time interval ratio at the ith pulse of the nth segment and $\overline{R_{SD,n}}$ is the mean systolic and diastolic time interval ratio of the nth segment. The $R_{SD,n,i}$ is calculated by

$$R_{SD,n,i} = \left(T_{trough,n,i} - T_{peak,n,i}\right)/\left(T_{peak,n,i} - T_{trough,n-1,i}\right) \quad (2)$$

where $T_{trough,n,i}$ denotes the trough times (or lowest point) at the ith pulse of the nth segment.

Mean-Standard Deviation of Pulse Shape (STD_{WAV})

To derive pulse shape, we take N_{samp} sample points of a pulse. The $STD_{WAV,n}$ of the nth segment is derived by taking average of the standard deviation at each sample point as follows:

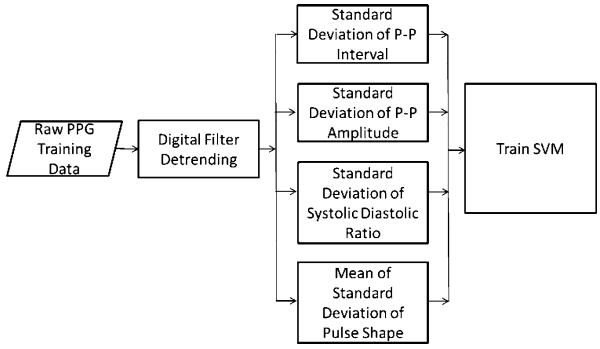


FIGURE 2. Training phase of the proposed SVM-based motion detection algorithm. Four time-domain features corresponding to (1) standard deviation of peak-to-peak intervals (2) standard deviation of peak-to-peak amplitudes (3) standard deviation of systolic and diastolic interval ratio, and (4) mean standard deviation of pulse shape, are candidate input variables to the SVM.

$$STD_{WAV,n} = \mathrm{E}[STD_{WAV,n,m}] \quad (3)$$

where $STD_{WAV,n,m}$ is calculated by substituting $D_{n,i}$ and $\overline{D_n}$ in (Eq. 1) with $q_{n,i}(m)$ and $\overline{q_n(m)}$, respectively, where $q_{n,i}(m)$ is the mth sample at the ith pulse of the nth segment and $\overline{q_n(m)}$ the mean at the mth sample of the nth segment.

SVM-BASED DETECTION OF MOTION/NOISE ARTIFACTS

Classification by Support Vector Machine (SVM)

SVM was applied to build a decision boundary classifying motion corruption from clean PPG signals. SVM is widely used in classification and regression due to its accuracy and robustness to noise.[25] The SVM consists of training and test phases described in the following sections.

Training Phase

A flow chart of the training phase in the SVM-based MNA detection algorithm is shown in Fig. 2. The SVM first derives the parameter values from clean and corrupted PPG training segments which are labeled separately (clean: 0, corrupted: 1). The SVM then trains itself with the labeled parameter values and finds the support vectors among the parameter values which maximize the margin (or the distance) between different classes. Finally, the SVM builds a decision boundary from the support vectors. If the estimated decision from the decision boundary is different from its known label, the decision is regarded as a *training error*. We consider a soft-margin SVM which can set the boundary even when the data sets are mixed and cannot be separated. In the soft-margin SVM

algorithm, slack variables are introduced to minimize the training error while maximizing the margin. Soft-margin SVM uses the following equation to find the support vectors.[13]

$$\text{Minimize } C\sum_{sv=1}^{N}\delta_{sv}+\frac{1}{2}\langle \mathbf{w}_s,\mathbf{w}_s\rangle,$$

$$\text{Subject to } T_{sv}(\langle \mathbf{w}_s,\mathbf{y}_{sv}\rangle+b_s)\geq 1=\delta_{sv} \text{ for } sv=1,2,\ldots,N \ =1,2,\ldots,N, \text{ and } \delta_{sv}\geq 0 \quad (4)$$

where C is regulation parameter, N is the number of vectors, δ_{sv} is the slack variable, $\mathbf{w}_s$is the weight vector and $<\cdot,\cdot>$ is the inner product operation. The $T_{sv}$is the svth target variable, $\mathbf{y}_{sv}$ are the svth input vector data, and b_s is the bias. The SVM decision boundary F_{sv} is derived as

$$F_{sv}=\langle \mathbf{w}_s^*,\mathbf{y}\rangle+b_s^*=0 \quad (5)$$

where $\mathbf{w}_s^*$ and b_s^* are weight factor and bias, respectively, obtained from Eq. (4), and $\mathbf{y}$ is the input point.

By transforming the $\mathbf{y}_{sv}$ and $\mathbf{y}$ term to $\mathbf{y}_{sv}\rightarrow(\mathbf{y}_{sv})$ and $\mathbf{y}\rightarrow(\mathbf{y})$, the non-linear SVM can be transformed to a linear SVM. For nonlinear SVM, Eq. (4) is modified as

$$T_{sv}(\langle \mathbf{w}_s,(\mathbf{y}_{sv})\rangle+b_s)\geq 1 \quad (6)$$

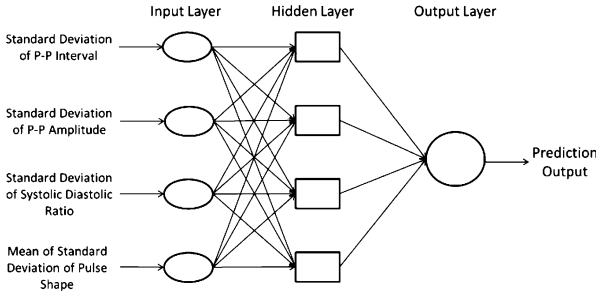


FIGURE 3. Test phase of the proposed SVM-based motion detection algorithm. The hidden layers correspond to kernel function of the SVM. The function between hidden layer and output layer is a linear operator.

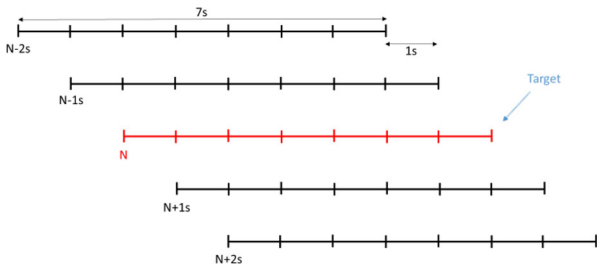


FIGURE 4. Enhancement of MNA detection by diversity. Neighbor segments are the segments surrounding a target segment within ±2 s. Decisions on the target segment are based on a majority vote from the decisions of neighbor segments as well as the one of the target segment (red).

To facilitate the operation on nonlinear SVM, a kernel function $K_s(\cdot,\cdot)$, which is a dot-product in the transformed feature space as follows, is used,

$$K_s(\mathbf{y}_{sv},\mathbf{y}_{sv'})=\langle(\mathbf{y}_{sv}),(\mathbf{y}_{sv'})\rangle \quad (7)$$

where $sv'=1,2,\ldots,N$.

Test Phase

Figure 3 shows a flow chart of the test phase in the SVM-based MNA detection algorithm. We partition PPG data into many 7-s segments, and derive parameters from each PPG portion to examine if it is corrupted by motion artifacts or not.

Enhancement of MNA Detection by Major Votes

To enhance MNA detection performance, the proposed algorithm incorporates multiple decisions on a set of neighbor segments in deciding whether a "target" segment is clean or corrupted. A neighbor segment is defined as a segment surrounding a target segment within $\pm T_{\text{neighbor}}$ seconds. A decision on a neighbor segment is highly likely to be the same as the decision on a target segment since PPG pulses in the neighbor segments are most likely to exhibit similar dynamics to the target segment.

The algorithm gathers the decisions of neighbor segments as well as target segment (see Fig. 4) and makes a final decision on the target segment based on a majority vote concept.

RESULTS

We evaluated the performance of the MNA detection algorithm for various types (simulated, laboratory controlled, and daily activities) of motion-corrupted PPGs to validate its performance in a wide range of scenarios. For all types of motions, the PPG recordings were divided into 7-s segments since this was determined to be the optimal size among the data length tested from 3 to 11 s (see Section IV-B). Another study[41] has also found the optimal segment size to be 7-s, hence, this allowed direct comparison between the two algorithms. We compared the proposed algorithm with four recently published MNA detection algorithms based on kurtosis (K), Shannon entropy (SE), Hjorth 1 (H1), and Hjorth 2 (H2) metrics,[9,34,41] respectively. As performance metrics, we considered classification accuracy, sensitivity and specificity. We also investigated mean HR and SpO_2 errors as well as detection error ratio.

Reference: Clean vs. corrupted

The following are criteria which we adopted to reference PPG segments (clean or corrupted) for each experiment. A visual reference was excluded to avoid subjective decisions by visual inspectors; for subtle MNA there were large disagreements among visual inspectors. Instead, we performed objective decisions based on controlled corruption start ($T_{corr,start}$) and end ($T_{corr,end}$) time points, ECG-derived heart rate (HR_{ECG}), PPG-derived heart rate (HR_{PPG}), and SpO_2 (SpO_{2PPG}) from PPG signals. HRs from ECG signals and pulse rates (PRs) from PPG signals are shown to have high agreement when the subjects do not move.[3] Specifically, the difference between HRs and PRs for 10 subjects is within approximately 5 beats/min on average. Even during heavy exercise, sudden increases or decreases in HR between successive beats are within 20 beats/min[6] while sudden SpO_2 increases or decreases between successive beats are around 2%.[1] Since subjects' movements in this work are less severe than heavy exercise, we expect sudden HR and SpO_2 changes to be smaller than the numbers noted above. Based on these experimental observations, we set the clean vs. motion-corrupted data classification criteria for PPG segments as follows.

Laboratory controlled data (forehead and finger)

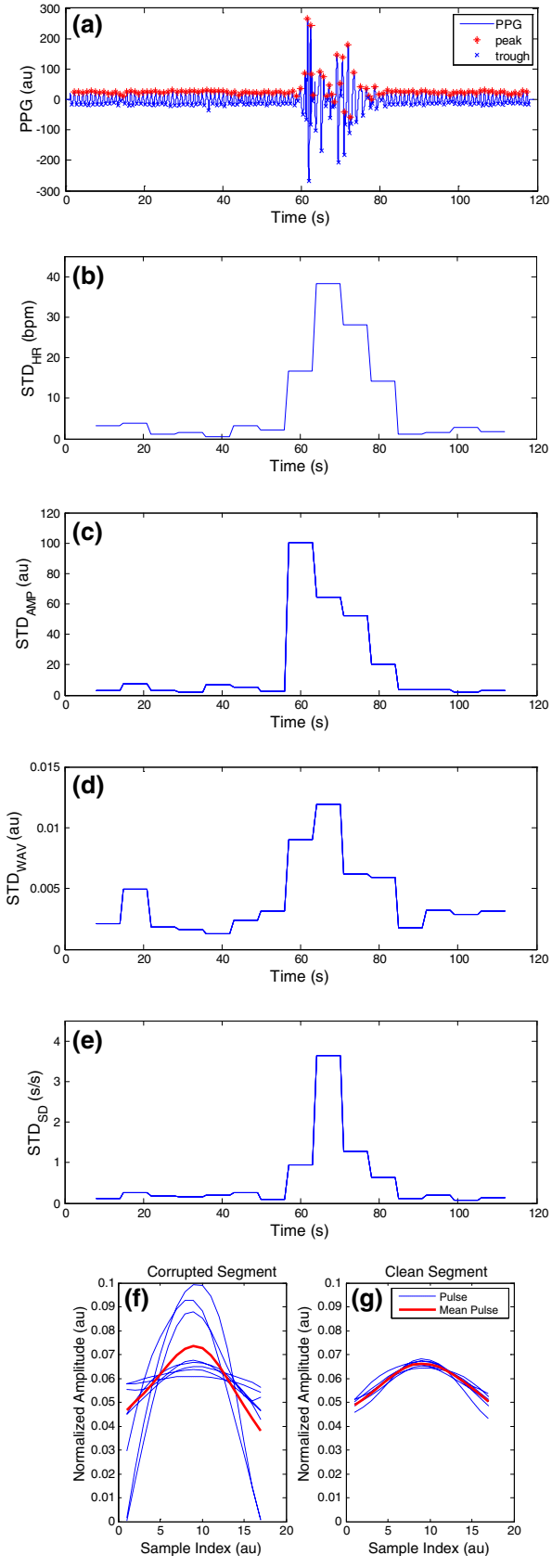
- If more than 85% of a segment is outside of [$T_{corr,start}$, $T_{corr,end}$], the segment was considered *clean*. Otherwise, the segment was referenced to be *corrupted*.
- If SpO_{2PPG} deviates by 10% from the mean of SpO_{2PPG} in a segment, then the segment was referenced to be corrupted.
- Successive difference, $|\text{diff}(HR_{PPG}(i+1) - HR_{PPG}(i))|$, from PPG signals is larger than 20 bpm for at least one pulse during a segment, then the segment was referenced to be corrupted.

Daily activity data (walking and stair-climbing)

- Successive difference, $|\text{diff}(HR_{ECG}(i+1) - HR_{ECG}(i))|$, from ECG signals is larger than

TABLE 1. Numbers of subjects and numbers of clean and corrupted segments per each motion artifact.

Type	Subtype	# of subjects	# of clean	# of corrupted
Simulation	Simulation	N/A	N/A	N/A
Laboratory controlled	Finger	13	195	105
	Forehead	11	190	110
Daily-activity	Walking/ stair-climbing	9	125	175



◀ **FIGURE 5. A sample forehead recorded PPG signal (a) along with the (b) standard deviation of *P–P* intervals (c) standard deviation of *P–P* amplitudes (d) standard deviation of systolic and diastolic time ratio, and (e) mean standard deviation of pulse shape, computed for each segment. The normalized sampled corrupt and clean PPGs for mean standard deviation of pulse shape (N_{samp} = 17) are given in (f).**

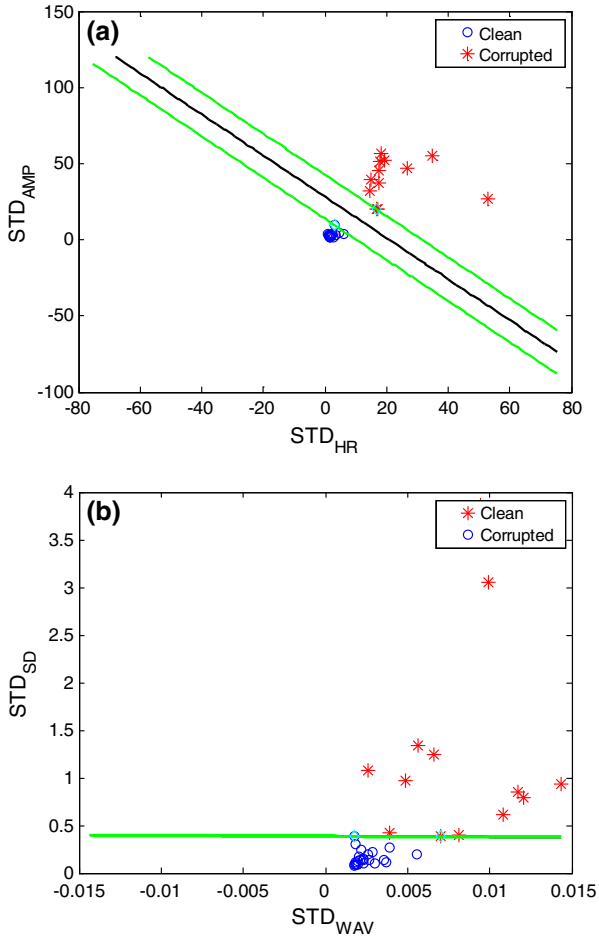


FIGURE 6. Trained SVM classification with a sample training finger recorded PPG signal is given with (a)–(b) pairs of two parameters. The SVM decision and margin boundaries are marked by black and green lines, respectively.

20 bpm for at least one pulse during a segment, then the segment was excluded.

- If SpO_{2PPG} deviates by 10% from the mean of SpO_{2PPG} in a segment, then the segment was referenced to be corrupted.
- If $|\text{diff}(HR_{PPG}(i+1) - HR_{PPG}(i))|$ is larger than 20 bpm for at least one pulse during a segment, then the segment was referenced to be corrupted.
- If $|HR_{ECG} - HR_{PPG}| < 5$ bpm during more than 85% of a segment, the segment was considered clean. Otherwise, the segment was referenced to be corrupted.

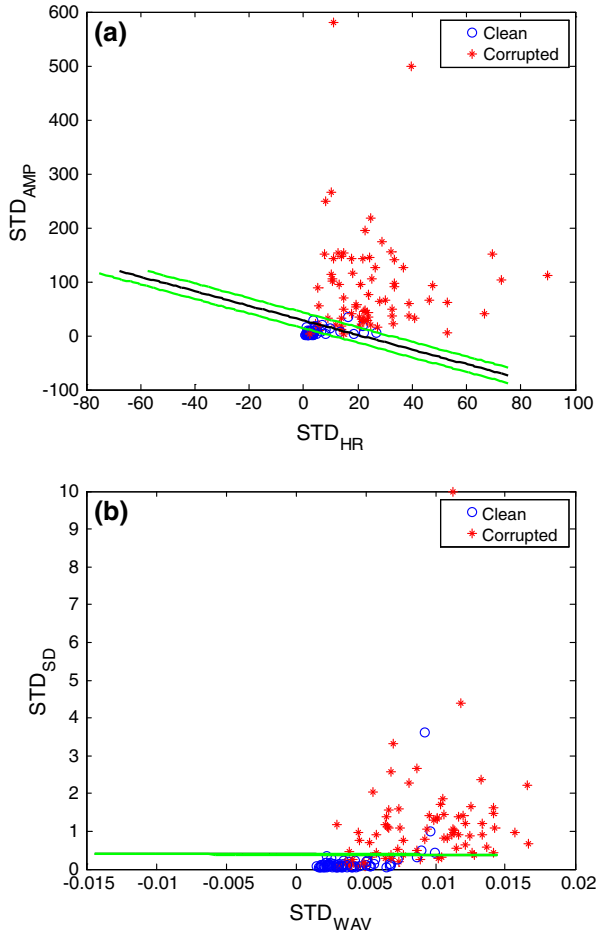


FIGURE 7. Validation: pairs of parameters for clean and corrupted PPG signals.

Table 1 describes the number of clean and corrupted PPG segments for each motion type used in the experiment as determined by the criteria defined above

Classification Accuracy

A sample forehead PPG signal and its corresponding 56–85 s have larger parameter values compared to clean segments between 1–56 s and 85–112 s. We sampled 17 points ($N_{samp} = 17$) of each pulse using the spline interpolation irrespective of test subjects and conditions to derive pulse shapes as shown in Fig. 5f. This enabled pulse shape comparisons within and among test subjects.

Figures 6a and 6b show $(STD_{HR}, STD_{AMP})$and (STD_{SD}, STD_{WAV}) of clean (circle) and corrupted (star) forehead signals, respectively, with corresponding SVM boundaries (black line). To lower computational complexity, a linear kernel was considered for the SVM in the experiment. We adopted inverse *k-fold* cross-validation which selects one-fold for training and the remaining k – 1-folds for testing.[33] We trained

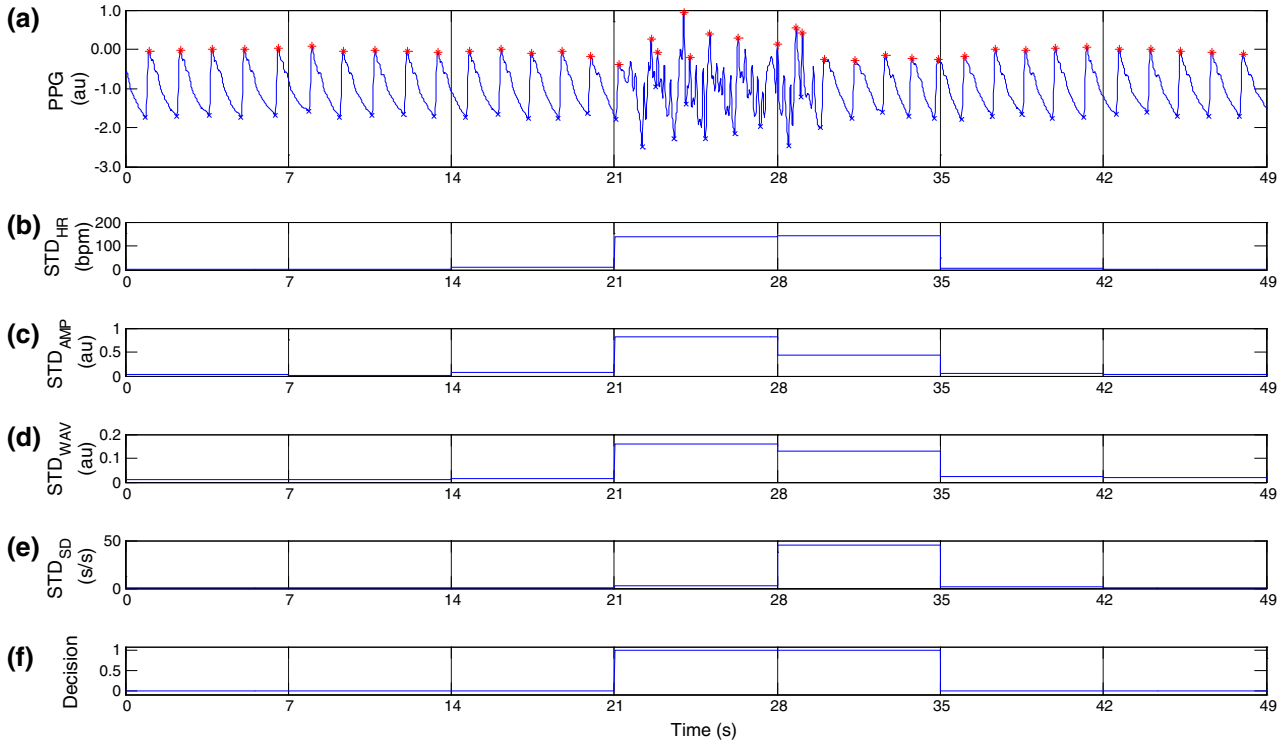


FIGURE 8. A representative PPG signal with detected peaks (red) (a) along with the (b) standard deviation of *P–P* intervals (c) standard deviation of *P–P* amplitudes (d) mean standard deviation of pulse shape and (e) standard deviation of systolic and diastolic time ratio, computed for each segment.

SVM with data from one subject after removing outliers and tested with data from 10 subjects. While there were small differences in the test errors among different k-fold cross-validation cases, we selected the least test-error case as the training data for each subject. Retraining was not performed since each subject's data consisted of a sufficient number of clean and corrupted PPG signals and the signal-to-noise ratios of the clean PPG signals were usually higher than the algorithm's sensitivity level. We optimized regularization parameter value (C) of the linear kernel SVM in terms of minimizing the training error rate. We adopted a 11-fold cross-validation and grid search ($C = \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3}\}$) which is widely used to determine C.[4] Figure 7 shows classification results by the SVM boundaries obtained from Fig. 6. Figure 8 shows a representative PPG signal with detected peaks (red) along with the corresponding statistical parameter values. Note the corrupted PPG signal interval between 21 to 31 s. The discrepancy between corrupted and clean portions is reflected by parameters STD_{HR}, STD_{AMP}, STD_{SD} and STD_{WAV}. The parameter values from the corrupted PPG segments exhibit larger variability and consequently have higher standard deviation value compared to those from clean data segments. The STD_{HR}, STD_{AMP} and STD_{WAV} have large values between 21 and 35 s (see Figs. 8b, 8c, and 8d), while STD_{SD} has large value only between 21 and 28 s (see Fig. 8e). Using SVM with these parameter values, the proposed algorithm correctly discriminated MNA corrupted segment between 21 and 35 s (see Fig. 8f). Table 2 presents C for finger, forehead, and walking/stair-climbing data. We tested our algorithm to different segment lengths varying from 3 to 11 s and calculated their mean classification accuracies, which are provided in Table 3. Among the different data segment lengths tested, the 7-s segment provided the highest classification accuracies for all data: finger, forehead and walking/stair-climbing PPG signals. Accuracy, specificity, and sensitivity for each dataset are presented in Table 4. On average, the SVM performance using the 7-s segment showed a 93.9% accuracy, 92.4% specificity, and 94.3% sensitivity. To derive the lower bound of our algorithm's

TABLE 2. *C* obtained by ninefold cross-validation and grid search method.

Type	Subtype	*C*
Simulation	Simulation	100
Laboratory controlled	Finger	1000
	Forehead	1
Daily-activity	Walking/stair-climbing	0.01

TABLE 3. Detection accuracy (mean ± SD) for varying window length.

	Window length				
Type	3	5	7	9	11
Finger	0.883 ± 0.042	0.906 ± 0.035	0.944 ± 0.033	0.944 ± 0.033	0.875 ± 0.042
Forehead	0.883 ± 0.023	0.880 ± 0.027	0.934 ± 0.035	0.856 ± 0.045	0.805 ± 0.044
Walk-stair climbing	0.813 ± 0.033	0.871 ± 0.039	0.937 ± 0.026	0.867 ± 0.052	0.856 ± 0.056

TABLE 4. Detection accuracy, specificity and sensitivity (mean ± SD) for 7-s segment.

Type	Accuracy	Specificity	Sensitivity
Finger	94.4 ± 3.3	94.7 ± 4.5	94.7 ± 3.4
Forehead	93.4 ± 3.5	96.7 ± 3.0	88.8 ± 7.9
Walking/stair-climbing	93.7 ± 2.7	91.4 ± 2.0	93.9 ± 5.0

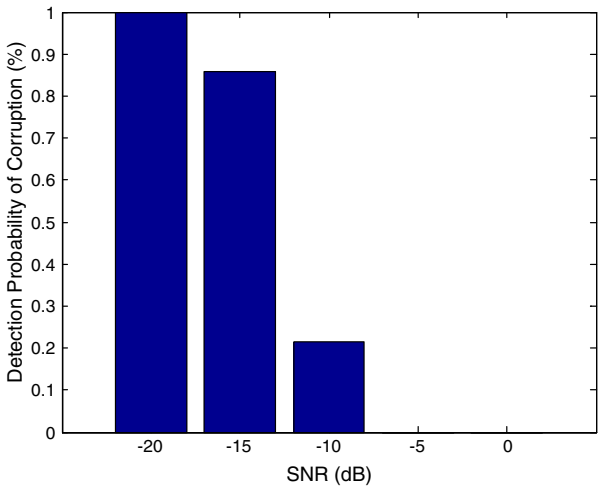


FIGURE 9. Detection probability of corruption by additive white Gaussian noise (AWGN) for varying SNR from −20 to 0 dB. 50 AWGN realizations for each SNR level are separately added to a non-MNA corrupted PPG. Each realization is tested by the proposed MNA detection algorithm to compute the detection probability of corruption.

performance[8] as well as to evaluate the sensitivity of our MNA detection algorithm, we added Gaussian white noise (GWN) of varying signal-to-noise (SNR) levels to a representative non-MNA corrupted PPG signal. For each SNR, 50 independent realizations of clean PPG signal with GWN are generated. As shown in Fig. 9, the PPG signals with a SNR below −10 dB are detected as corrupted data with our algorithm. For a SNR of −20 dB, every segment was detected as corrupted.

Performance Comparison of MNA detection Algorithms

Our algorithm was compared with other artifact detection methods based on H1, H2, K and SE since these methods have been shown to provide good detection accuracies.[9–11] The H1 and H2 parameters represent the central frequency and half of bandwidth, respectively, and are defined as follows:

$$H_1 = \sqrt{\frac{\overline{v}_2(n)}{\overline{v}_0(n)}} \text{ and } H_2 = \sqrt{\frac{\overline{v}_4(n)}{\overline{v}_2(n)} - \frac{\overline{v}_2(n)}{\overline{v}_0(n)}} \quad (8)$$

where $\overline{v}_i(n) = \int_{-\pi}^{\pi} v^i S_{ypDC}(e^{Jv})dv$. Here, $S_{ypDC}(e^{Jv})$ is the power spectrum of signal $y_{pDC}(n)$.

For a fair comparison, all detection methods used 7 s data segments. Figures 10a, 10b, and 10c compare the medians and 25th and 75th percentiles of detection accuracy, sensitivity, and specificity, which are obtained from each subject, for all five detection methods for the finger, head and walking/stair-climbing data sets. In general, our SVM method consistently yielded higher performance with a mean accuracy of 94%, sensitivity of 97%, and a specificity of 92%; whereas other methods showed fluctuations depending on which datasets were used. In the finger recorded data, H1 yielded a slightly higher accuracy than all other methods due to higher specificity, but the detection sensitivity was lower.

HR and SpO_2 Estimation

Figure 11a compares the mean HR error and detection error fraction from five MNA detection methods for walking/stair-climbing data. The HR errors were defined by the difference between the estimated HR derived from the PPG and the reference HR readings; the detection error fraction is defined as a ratio of the number of erroneous detection events to that of total detection events. Low mean HR error and low detection error fraction would reflect an effective artifact detection algorithm. Our algorithm yielded the lowest HR error and detection error fraction among

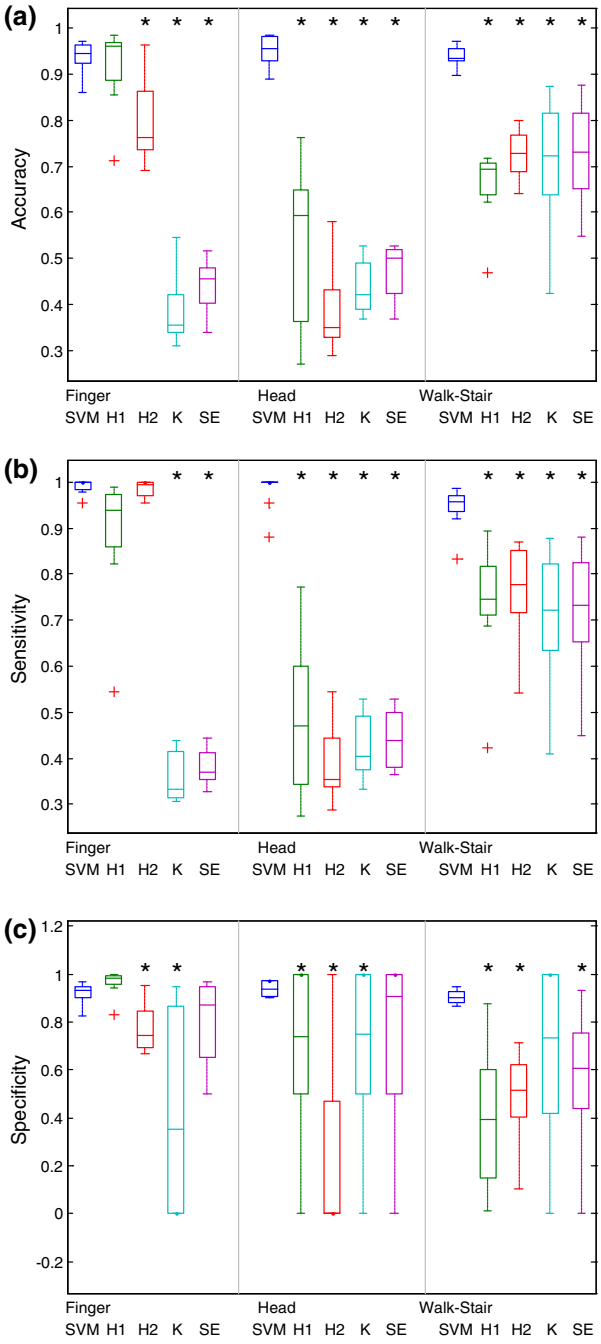


FIGURE 10. Classification performance comparison between our SVM algorithm, Hjorth (H1, H2), Kurtorsis and Shanon Entropy (K, SE) parameters. (a) Accuracy; (b) Sensitivity; (c) Specificity. The central mark on each box corresponds to the median; the edges of the box correspond to the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. (*) indicate the mean is significantly different ($p < 0.05$ at 95% CI) between SVM and other methods used for comparison.

the other MNA methods we compared. Figure 11b compares mean SpO_2 error and detection error fraction from five MNA detection methods. The SE based detection method showed a lower mean SpO_2 error than our algorithm, but its detection error fraction was very high (>70%), indicating that the error was computed based on only 30% of clean data. On the other hand, the proposed SVM algorithm resulted in a mean SpO_2 error of 2.7 with a detection error of only 6.3%. Figure 12 compares five MNA detection methods in terms of paired t test results of HR and SpO_2 estimation and detection accuracy. On average, the SVM algorithm outperformed the K, SE, H1 and H2 methods with HR errors of 2.3 bpm, SpO_2 errors of 2.7% and detection error fraction of 6.3%.

DISCUSSION

Robust real-time MNA detection algorithms for raw PPG signals have been elusive to date. In this study, an SVM-based method is introduced to detect MNA-corrupted PPG data. Reconstruction of MNA-corrupted PPG segments is described in the companion paper. The aim of the current paper is to detect the MNA-corrupted PPG segments as accurately as possible. The question is how to detect MNA in an adaptive way to maximize detection accuracy so that PPG signal distortions is minimized. To address this question, we used four parameters derived from the PPG data: (a) standard deviation of peak-to-peak intervals (b) standard deviation of peak-to-peak amplitudes (c) standard deviation of systolic and diastolic time ratios, and (d) mean-standard deviation of pulse shapes. These four parameters are then used as inputs to an SVM-based MNA detection method with a sliding window with a major vote concept. We use these parameters since motion corrupted PPG signals are observed to have noticeably different amplitude values and have large variations for successive pulses when compared to the clean PPG signals.[35] We are currently working on finding more PPG relevant parameters to enhance detection performance on diverse types of MNA-corrupted PPG signals.

Many detection algorithms have been proposed to detect MNAs or quantify signal quality of PPG pulses. Various approaches use a set of several PPG-derived parameters to detect MNA, but the test data was confined to limited types of motions.[7,14,17,21,22,35,41] Given that most methods provide adequate MNA detection performance, we evaluated our proposed algorithm by comparing it to them.

The results demonstrated that our proposed SVM-based MNA detection algorithm provided higher classification accuracy as well as lower HR and SpO_2 errors compared to the conventional detection methods. The paired t test was performed to determine if there is significant difference between classification

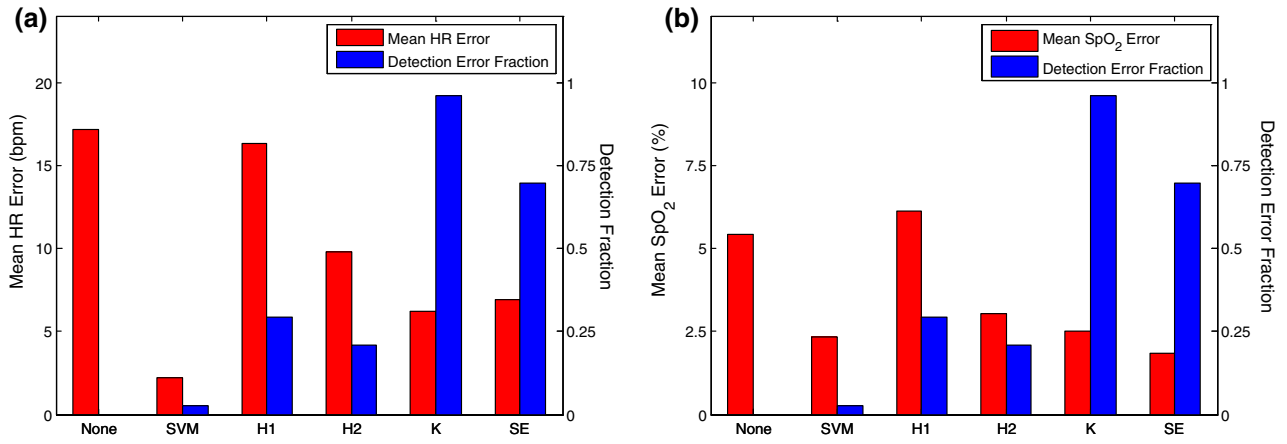


FIGURE 11. Comparison of mean errors and detection error fraction between original signal (labeled "None") and artifact removed signal from five detection methods (SVM, H1, H2, K, and SE). (a) HR error; (b) SpO_2 error.

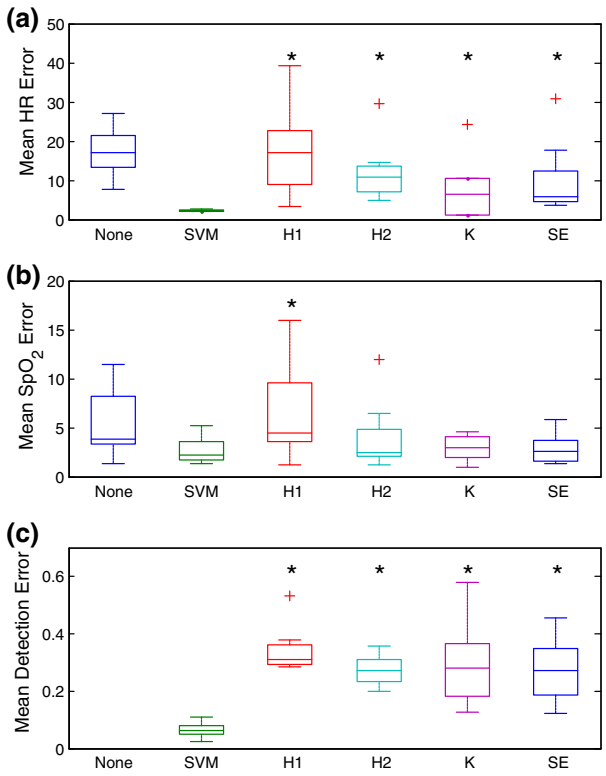


FIGURE 12. Mean error comparison between our SVM algorithm, Hjorth (H1, H2), Kurtorsis and Shanon Entropy (K, SE) parameters. (a) heart rate; (b) SpO_2; (c) detection error for walk/stair-climbing data. The central mark on each box corresponds to the median; the edges of the box correspond to the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. (*) indicate the mean is significantly different ($p < 0.05$ at 95% CI) between SVM and other methods used for comparison. The *X*-axis labelled "None" in all panels refers to the mean errors when compared to the reference signals without removing the MNA detected segments as identified by any of the five computational methods.

errors obtained from our SVM approach compared with other published methods. For the finger recorded PPG segments, Fig. 10a indicates that the mean classification accuracy is significantly different ($p < 0.05$ at 95% CI) between our SVM and K, SE and H2 methods (except for H1). K, SE, H1 and H2 methods were significantly different from our SVM method for forehead and walking/stair-climbing PPG data. Kurtosis and Shannon entropy based detection methods' low performance may be caused by clean PPG signals' variations in amplitude and pulse. These variations can be induced unintentionally during the measurement due to minimal movement or physiological artifacts which are not severe but vary between test subjects or test conditions. Since H1 is based on estimating central frequency, the H1 method is expected to have high performance as the clean PPG signal gets more stable in the frequency domain. Hence, H1 shows high performance for finger movement when clean PPG signals are measured in the most stable conditions for finger, head and walking/stair-climbing PPG measurements as shown in Fig. 10a, 10b, and 10c. H1 performs the worst for head and walking/stair-climbing PPG data when slight movement and physiological artifacts become more prominent. Similarly, the H2 method which estimates the half of the bandwidth of H1 shows comparable performance for finger PPG signals as shown in Fig. 10a, 10b, and 10c. However, H2 performs worse for head and walking/stair-climbing PPGs. Figure 12a, 12b, and 12c summarizes paired-*t* test results of HR and SpO_2 estimations as well as detection accuracy for walking/stair-climbing data. As shown in Fig. 12a, 12b, and 12c, SVM is significantly different from H1, H2, K, and SE in terms of HR estimation and detection accuracy (see Figs. 11a and

11c), while SpO_2 derived from the SVM method is significantly different from only H1 (see Fig. 12b).

To enhance the decision accuracy of our MNA detection algorithm, we adopted a major vote concept which is widely used in other engineering fields to fuse the decisions from multiple entities.[26] The major vote concept in our algorithm took a role in providing a decision on a target PPG segment after fusing the decisions of neighbor PPG segments as well as that of the target segment. This is based on the observation that the decision on the target segment (clean or corrupted) is highly correlated to those of neighbor segments.

The advantage of our MNA detection algorithm is that it can classify MNA-corrupted PPG from clean PPG in an adaptive manner. Hence, it can be applied to either controlled or daily-activity moving scenarios. The MNA detection algorithm coded with Matlab (2012a) takes only 7 ms on an Intel Xeon 3.6 GHz computer for the 7-s data segment. Hence, the algorithm is real-time realizable especially when it is coded in either C or C++. Our proposed MNA algorithm, when incorporated with the reconstruction algorithm as detailed in our companion paper (ref), can also provide information on whether the signal is clean, relatively noise-free after reconstruction (detected as corrupted but succeeded in reconstruction) or MNA-corrupted (detected as noisy and failed in reconstruction). In conclusion, the potential for the method proposed in this work to have practical applications is high, and the integration of the algorithm described with a pulse oximeter device may have significant implications for real-time clinical applications and especially for ambulatory or smart phone monitoring of vital signs.

ACKNOWLEDGMENTS

This work was supported in part by the US Army Medical Research and Materiel Command (USAMRMC) under Grant No. W81XWH-12-1-0541.

REFERENCES

[1]Andersen, P., and B. Saltin. Maximal perfusion of skeletal muscle in man. *J. Physiol.* 366:233–249, 1985.

[2]Barker, S. J., and N. K. Shah. The effects of motion on the performance of pulse oximeters in volunteers (revised publication). *Anesthesiology* 86:101–108, 1997.

[3]Chang, K.-M., and K.-M. Chang. Pulse rate derivation and its correlation with heart rate. *J. Med. Biol. Eng.* 29:132–137, 2009.

[4]Comtois, G., Y. Mendelson, and P. Ramuka. A comparative evaluation of adaptive noise cancellation algorithms for minimizing motion artifacts in a forehead-mounted wearable pulse oximeter. Engineering in Medicine and Biology Society, 2007 EMBS 2007 29th Annual International Conference of the IEEE, 2007, pp. 1528–1531.

[5]Dash, S., K. H. Chon, S. Lu, *et al.* Automatic real time detection of atrial fibrillation. *Ann. Biomed. Eng.* 37:1701–1709, 2009.

[6]Engelen, M., J. Porszasz, M. Riley, *et al.* Effects of hypoxic hypoxia on O_2 uptake and heart rate kinetics during heavy exercise. *J. Appl. Physiol.* 81:2500–2508, 1996.

[7]Foo, J. Y., and S. J. Wilson. A computational system to optimise noise rejection in photoplethysmography signals during motion or poor perfusion states. *Med. Biol. Eng. Comput.* 44:140–145, 2006.

[8]Ganeshapillai, G., and J. Guttag. Real time reconstruction of quasiperiodic multi parameter physiological signals. *EURASIP J. Adv. Signal Process.* 1–15:2012, 2012.

[9]Gil, E., J. María Vergara, and P. Laguna. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. *Biomed. Signal Process. Control* 3:267–277, 2008.

[10]Hjorth, Bo. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29:306–310, 1970.

[11]Hjorth, Bo. The physical significance of time domain descriptors in EEG analysis. *Electroencephalogr. Clin. Neurophysiol.* 34:321–325, 1973.

[12]Hong Enriquez, R., M. Sautie Castellanos, J. Falcon Rodriguez, *et al.* Analysis of the photoplethysmographic signal by means of the decomposition in principal components. *Physiol. Meas.* 23:N17–N29, 2002.

[13]Hsu, C.-W., C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Tech. rep., Department of Computer Science, National Taiwan University, 2003.

[14]Karlen, W., K. Kobayashi, J. M. Ansermino, *et al.* Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation. *Physiol. Meas.* 33:1617–1629, 2012.

[15]Kim, B. S., and S. K. Yoo. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng.* 53:566–568, 2006.

[16]Krishnan, R., B. Natarajan, and S. Warren. Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data. *IEEE Trans. Biomed. Eng.* 57:1867–1876, 2010.

[17]Lee, B., J. Han, H. J. Baek, *et al.* Improved elimination of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry. *Physiol. Meas.* 31:1585–1603, 2010.

[18]Lee, H. The periodic moving average filter for removing motion artifacts from PPG signals. *Int. J. Control Autom. Syst.* 5:701–706, 2007.

[19]Lee, J., W. Jung, I. Kang, *et al.* Design of filter to reject motion artifact of pulse oximetry. *Comput. Stand. Interfaces* 26:241–249, 2004.

[20]Li, K., and S. Warren. A wireless reflectance pulse oximeter with digital baseline control for unfiltered photoplethysmograms. *IEEE Trans. Biomed. Circuits Syst.* 6:269–278, 2012.

[21]Li, K., S. Warren, and B. Natarajan. Onboard tagging for real-time quality assessment of photoplethysmograms

acquired by a wireless reflectance pulse oximeter. *IEEE Trans. Biomed. Circuits Syst.* 6:54–63, 2012.
[22]Li, Q., 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.* 29:15–32, 2008.
[23]Mannacio, V., L. Di Tommaso, V. De Amicis, *et al.* Coronary perfusion: impact of flow dynamics and geometric design of 2 different aortic prostheses of similar size. *J. Thoracic Cardiovasc. Surg.* 143:1030–1035, 2012.
[24]Nakajima, K., T. Tamura, and H. Miike. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Med. Eng. Phys.* 18:365–372, 1996.
[25]Naraharisetti, K. V. P., M. Bawa, and M. Tahernezhadi. Comparison of different signal processing methods for reducing artifacts from photoplethysmograph signal. 2011 IEEE International Conference on Electro/Information Technology (EIT), 2011, pp. 1–8.
[26]Olfati-Saber, R., J. S. Shamma. Consensus filters for sensor networks and distributed sensor fusion. 2005 and 2005 European Control Conference CDC-ECC'05 44th IEEE Conference on Decision and Control, 2005, pp. 6698–6703.
[27]Patterson, J. A. C., and Y. Guang-Zhong. Ratiometric artifact reduction in low power reflective photoplethysmography. *IEEE Trans. Biomed. Circuits Syst.* 5:330–338, 2011.
[28]Petterson, M. T., V. L. Begnoche, and J. M. Graybeal. The effect of motion on pulse oximetry and its clinical significance. *Anesth. Analg.* 105:S78–S84, 2007.
[29]Ram, M. R., K. V. Madhav, E. H. Krishna, *et al.* A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Trans. Instrum. Meas.* 61:1445–1457, 2012.
[30]Ram, M. R., K. V. Madhav, E. H. Krishna *et al.* Use of spectral estimation methods for computation of SpO_2 from artifact reduced PPG signals. Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE, 2011, pp. 431–436.
[31]Rusch, T. L., R. Sankar, and J. E. Scharf. Signal processing methods for pulse oximetry. *Comput. Biol. Med.* 26:143–159, 1996.
[32]Sahni, R., A. Gupta, K. Ohira-Kist, *et al.* Motion resistant pulse oximetry in neonates. *Arch. Dis. Child. Fetal Neonatal Ed.* 88:F505–F508, 2003.
[33]Samorodov, Av. Application of a fuzzy integral for weak classifiers boosting. *Pattern Recognit. Image Anal.* 21:206–210, 2011.
[34]Selvaraj, N., Y. Mendelson, K. H. Shelley *et al.* Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011, pp. 4972–4975.
[35]Sukor, J. A., S. J. Redmond, and N. H. Lovell. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiol. Meas.* 32:369–384, 2011.
[36]Sweeney, K. T., T. E. Ward, and S. F. Mcloone. Artifact removal in physiological signals—practices and possibilities. *IEEE Trans. Information Technol. Biomed.* 16:488–500, 2012.
[37]Tobin, R. M., J. A. Pologe, and P. B. Batchelder. A characterization of motion affecting pulse oximetry in 350 patients. *Anesth. Analg.* 94:S54–S61, 2002.
[38]Trivedi, N. S., A. F. Ghouri, N. K. Shah, *et al.* Effects of motion, ambient light, and hypoperfusion on pulse oximeter function. *J. Clin. Anesth.* 9:179–183, 1997.
[39]Wijshoff, R. W., M. Mischi, J. Veen, *et al.* Reducing motion artifacts in photoplethysmograms by using relative sensor motion: phantom study. *J. Biomed. Opt.* 17:117007, 2012.
[40]Yan, Y. S., C. C. Poon, and Y. T. Zhang. Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner–Ville distribution. *J. Neuroeng. Rehabil.* 2:3, 2005.
[41]Yu, C., Z. Liu, T. Mckenna, *et al.* A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *J. Am. Med. Inform. Assoc.* 13:309–320, 2006.

Annals of Biomedical Engineering, Vol. 42, No. 4, April 2014 (© 2013) pp. 885–898
DOI: 10.1007/s10439-013-0944-x

BMES BIOMEDICAL ENGINEERING SOCIETY™ www.bmes.org

Respiratory Rate Estimation from the Built-in Cameras of Smartphones and Tablets

YUNYOUNG NAM,[1] JINSEOK LEE,[2] and KI H. CHON[1]

[1]Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA; and [2]Department of Biomedical Engineering, Wonkwang University School of Medicine, Iksan, Jeonbuk, Republic of Korea

(*Received 28 August 2013; accepted 14 November 2013; published online 23 November 2013*)

Associate Editor Tingrui Pan oversaw the review of this article.

Abstract—This paper presents a method for respiratory rate estimation using the camera of a smartphone, an MP3 player or a tablet. The iPhone 4S, iPad 2, iPod 5, and Galaxy S3 were used to estimate respiratory rates from the pulse signal derived from a finger placed on the camera lens of these devices. Prior to estimation of respiratory rates, we systematically investigated the optimal signal quality of these 4 devices by dividing the video camera's resolution into 12 different pixel regions. We also investigated the optimal signal quality among the red, green and blue color bands for each of these 12 pixel regions for all four devices. It was found that the green color band provided the best signal quality for all 4 devices and that the left half VGA pixel region was found to be the best choice only for iPhone 4S. For the other three devices, smaller 50 × 50 pixel regions were found to provide better or equally good signal quality than the larger pixel regions. Using the green signal and the optimal pixel regions derived from the four devices, we then investigated the suitability of the smartphones, the iPod 5 and the tablet for respiratory rate estimation using three different computational methods: the autoregressive (AR) model, variable-frequency complex demodulation (VFCDM), and continuous wavelet transform (CWT) approaches. Specifically, these time-varying spectral techniques were used to identify the frequency and amplitude modulations as they contain respiratory rate information. To evaluate the performance of the three computational methods and the pixel regions for the optimal signal quality, data were collected from 10 healthy subjects. It was found that the VFCDM method provided good estimates of breathing rates that were in the normal range (12–24 breaths/min). Both CWT and VFCDM methods provided reasonably good estimates for breathing rates that were higher than 26 breaths/min but their accuracy degraded concomitantly with increased respiratory rates. Overall, the VFCDM method provided the best results for accuracy (smaller median error), consistency (smaller interquartile range of the median value), and computational efficiency (less than 0.5 s on 1 min of data using a MATLAB implementation) to extract breathing rates that varied from 12 to 36 breaths/min. The AR method provided the least accurate respiratory rate estimation among the three methods. This work illustrates that both heart rates and normal breathing rates can be accurately derived from a video signal obtained from smartphones, an MP3 player and tablets with or without a flashlight.

Keywords—Respiratory rate estimation, Autoregressive model, Continuous wavelet transform, Variable frequency complex demodulation method, Smartphone, Tablet.

INTRODUCTION

Respiratory rate is an important indicator for early detection and diagnosis of potentially dangerous conditions such as sleep apnea,[24] sudden infant death syndrome,[18] cardiac arrest[3] and chronic obstructive pulmonary disease.[5] In addition, for some patients who undergo surgery, relative changes in respiratory rates are much greater than changes in heart rate or systolic blood pressure, thus, respiratory rates can be an important vital sign indicator.[21] Respiratory rate is most accurately measured using transthoracic impedance plethysmography,[1] nasal thermocouples[20] or capnography.[16] However, these methods all require expensive external sensors which may require donning a mask, nasal cannula or chest band sensors. More importantly, since these devices may disturb natural breathing and sleep positions, they are mostly applicable in constrained environments such as operating rooms and intensive care units.

Recently, photoplethysmography (PPG) has been widely considered for respiratory rate extraction due to its simplicity and non-invasive measurement capability.[11–13] The PPG signal contains components that are synchronous with respiratory and cardiac rhythms. Indeed, the

Address correspondence to Ki H. Chon, Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA. Electronic mail: kichon@wpi.edu

respiratory rhythm is modulated by frequency and/or amplitude of the cardiac rhythm. The occurrence of temporal variations of frequency and amplitude is characteristic of the respiratory sinus arrhythmia.[6] Thus, the respiratory rate can be obtained by detecting the presence of either amplitude modulation (AM) or frequency modulation (FM) in the PPG signal.[2]

Numerous advanced signal processing algorithms (both parametric and nonparametric approaches) have been applied to extract respiratory rates by looking for AM or FM signatures from a PPG signal.[2,19] For a parametric approach, the autoregressive (AR) model approach has been shown to provide relatively good respiratory rate estimation.[7–10] For nonparametric approaches, time–frequency spectrum (TFS) methods such as continuous wavelet transform (CWT) and variable frequency complex demodulation method (VFCDM) have also been shown to provide accurate respiratory rate estimation.[2,11–13]

To our knowledge, respiratory rate estimation using the camera of either a smartphone or a tablet has never been demonstrated nor discussed in the literature. We have recently demonstrated that a pulsatile signal (PS) that has similar dynamics to that of a PPG signal can be obtained from a smartphone's camera when a fingertip is pressed onto it.[4,19] Utilizing these PS derived from an iPhone, we have also shown that accurate detection of atrial fibrillation can be made.[17] Given these advances, the aims of this work were: (1) a systematic examination of the PS quality derived from a video camera from several measurement modalities including iPhone 4S, iPad 2, iPod 5, and Galaxy S3; and (2) to determine if accurate respiratory rates can be estimated directly from the PS of the different measurement modalities. The challenge here is that PPG signals are often sampled at greater than 100 Hz whereas most smartphones' video sampling rates are no more than 30 Hz. Since previous studies have shown good estimation of respiratory rates using the AR model, CWT, and VFCDM from a PPG signal, we also use these methods to compare the accuracy of breathing rates from PS obtained from various models of a smartphone, MP3 player (iPod 5) and a tablet.

METHODS

Data Collection

Data were collected on 10 healthy subjects on 2 separate occasions using 4 different devices: iPhone 4S, iPad 2, iPod 5, and Galaxy S3. Only two devices were used simultaneously for data collection in a given experimental setting. Worcester Polytechnic Institute's Institutional Review Board approved the data collection technique. For the PS acquisition, we used the Objective-C programming language and the Xcode platform for iPhone 4S, iPad 2, and iPod 5; Java was used for the Galaxy S3 on the mobile platform Android 4.1 (Jelly Bean). Specifically, we used Eclipse IDE Indigo R2 for the development environment and Samsung Galaxy S3 for the development and debugging purposes. For the video recordings of iPhone, iPad, and iPod, we examined four different sizes of pixel regions: 50×50, 320×240 (QVGA), 640×240 (vertical HVGA), and 640×480 (VGA) for determining the optimal signal quality. For all five different pixel sizes, the PS was obtained by averaging the entire pixel size for each of the three color bands (red, green and blue) for every frame. All four devices provided sampling rate close to 30 frames per second. However, when the video sampling rate was lower than 30 Hz, a cubic spline algorithm was used to interpolate the signal to 30 Hz.

No subject had cardiorespiratory pathologies. All four devices were tested using the same subject, at the same location, and under the same test conditions. Data were collected in the sitting upright position, and the sensor was placed in proximity to the subject's left index or middle finger as shown in Fig. 1. All subjects were instructed to breathe at a metronome rate according to a timed beeping sound, i.e., to start inspiring when a beep sound starts and to expire before the next beep sound occurs. The data were collected for breathing frequencies ranging from 0.2 to 0.9 Hz at an increment of 0.1 Hz. Prior to data collection, all subjects were acclimated to the breathing frequency rate being measured. Three minutes of data were collected for each frequency for each subject. Electrocardiogram (ECG) recordings were collected with an HP 78354A acquisition system using a standard 5-lead electrode configuration. A respiration belt was placed around a subject's chest and abdomen to monitor the true breathing rate (Respitrace Systems, Ambulatory Monitoring Inc.). Respiratory and ECG recordings were obtained using the LabChart software (ADInstruments) at a sampling rate of 400 Hz. Figure 1 shows data collection on the four devices by placing a fingertip on the video camera.

Extraction of Respiratory Rates

VFCDM

Detection of AM and FM from a PS using the power spectral density (PSD) is difficult since the dynamics are time-varying, hence, require high resolution time–frequency spectral (TFS) methods to resolve them. We have recently shown that because the VFCDM method provides one of the highest TFS resolutions, it can identify AM and FM dynamics.

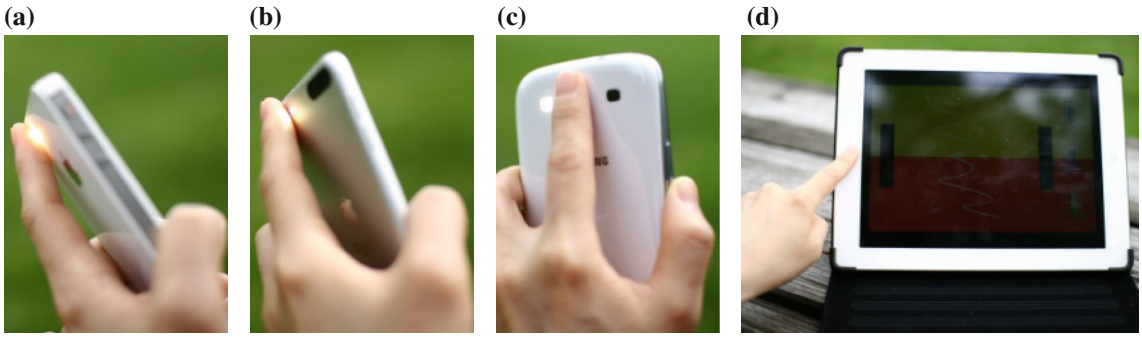


FIGURE 1. General scheme to acquire video from the four devices.

Consequently, Fourier transform of either the AM or FM time series extracted from the heart rate frequency band can lead to accurate estimation of respiratory rates when the acquired signal is PPG data.[23]

Details concerning the VFCDM algorithm are described in Wang *et al.*[23] Hence, we will only briefly describe the main essence of the algorithm. The VFCDM starts with an assumption that a signal $x(t)$ is considered to be a narrow band sinusoidal oscillation with a center frequency f_0, instantaneous amplitude $A(t)$, phase $\phi(t)$, and the direct current component $dc(t)$, as follows:

$$x(t) = dc(t) + A(t)\cos\left(2\pi f_0 t + \phi(t)\right) \tag{1}$$

For a given center frequency, instantaneous amplitude information $A(t)$ and phase information $\phi(t)$ can be extracted by multiplying Eq. (1) by $e^{-j2\pi f_0 t}$, resulting in the following:

$$z(t) = x(t)e^{-j2\pi f_0 t} = dc(t)e^{-j2\pi f_0 t} + \left(\frac{A(t)}{2}\right)e^{j\phi(t)} + \left(\frac{A(t)}{2}\right)e^{-j(4\pi f_0 t + \phi(t))}. \tag{2}$$

A leftward shift by $e^{-j2\pi f_0 t}$ results in moving the center frequency, f_0, to zero frequency in the spectrum of $z(t)$. If $z(t)$ in Eq. (2) is subjected to an ideal low pass filter (LPF) with a cutoff frequency $f_c < f_0$, then the filtered signal $z_{lp}(t)$ will contain only the component of interest and the following Eqs. (3a)–(3c) are obtained:

$$z_{lp}(t) = \left(\frac{A(t)}{2}\right)e^{j\phi(t)} \tag{3a}$$

$$A(t) = 2|z_{lp}(t)| \tag{3b}$$

$$\phi(t) = \arctan\left(\frac{\text{image}(z_{lp}(t))}{\text{real}(z_{lp}(t))}\right). \tag{3c}$$

When a modulating frequency is not fixed, as described above, but varies as a function of time, the signal $x(t)$ can be written in the following form:

$$x(t) = dc(t) + A(t)\cos\left(\int_0^t 2\pi f(\tau)d\tau + \phi(t)\right), \tag{4}$$

Similar to the operations in Eqs. (1) and (2), multiplying Eq. (4) by $e^{-j\int_0 2\pi f(\tau)d\tau}$ yields both instantaneous amplitude $A(t)$ and instantaneous phase $\phi(t)$, as described in the following equation:

$$z(t) = x(t)e^{-j\int_0^t 2\pi f(\tau)d\tau} = dc(t)e^{-j\int_0^t 2\pi f(\tau)d\tau} + \left(\frac{A(t)}{2}\right)e^{j\phi(t)} + \left(\frac{A(t)}{2}\right)e^{-j\left(\int_0^t 4\pi f(\tau)d\tau + \phi(t)\right)} \tag{5}$$

From Eq. (5), if $z(t)$is filtered with an ideal LPF with a cutoff frequency $f_c < f_0$, then the filtered signal $z_{lp}(t)$ will be obtained with the same instantaneous amplitude $A(t)$ and phase $\phi(t)$ as provided in Eqs. (3b) and (3c). The instantaneous frequency is given by:

$$f(t) = f_0 + \frac{1}{2\pi}\frac{d\phi(t)}{dt}. \tag{6}$$

The VFCDM method thus involves a two-step procedure. The first step is to use complex demodulation (CDM) or what we termed the fixed frequency CDM (FFCDM) to obtain an estimate of the TFS, and the second step is to select only the dominant frequencies of interest for further refinement of the time–frequency resolution using the VFCDM approach. In the first step of the VFCDM method, a bank of LPFs is used to decompose the signal into a series of band-limited signals. The analytic signals that are obtained from these, through use of the Hilbert transform, then provide estimates of the instantaneous amplitude, frequency, and phase within each frequency band.

CWT

As described in Introduction section, numerous studies[11–13] showed relatively good results using the CWT for extraction of respiratory rates directly from a

pulse oximeter. The Morlet wavelet was also applied to a half-length of five samples at the coarsest scale for estimating the scalogram of the PS.[22] The procedures of the CWT for extracting respiratory rates is nearly identical to the VFCDM in that identified AM and FM series are Fourier transformed to estimate respiratory rates.

AR Modeling

This approach involves estimation of AR model parameters using the optimal parameter search (OPS) criteria.[15] The AR parameters are formulated as the transfer function followed by factorization into pole terms. The real and complex conjugate poles define the power spectral peaks with the larger magnitude poles corresponding to higher magnitude peaks. The resonant frequency of each spectral peak is given by the phase angle of the corresponding pole. Among the poles, we set the region of interest for respiratory rates between 0.15 and 1 Hz. The details of the respiratory algorithm using the AR model are described in Lee and Chon.[7]

Data Analysis

Using PPG signals with sampling rates of at least 250 Hz to derive heart rates has previously been shown to be a good alternative to ECG monitoring.[14] However, sampling rates for most smart phone and tablet video cameras range from 25 to 30 Hz. Given these low sampling rates, it is necessary to determine the accuracy of the smart phone and tablet devices in estimating heart rates and respiratory rates. Comparisons of derived heart rates were made between the standard ECG, smartphones and tablets. We used our own peak detection algorithm to determine R-wave peaks from the ECG signals and cardiac pulse peaks from the phone camera PPG signal. Due to the frame rate variability, we interpolated the PS to 30 Hz using a cubic spline algorithm followed by the peak detection. The peak detection algorithm incorporated a filter bank with variable cutoff frequencies, spectral estimates of the heart rate, rank-order nonlinear filters and decision logic.

Three minutes of data sampled at 30 Hz were low-pass-filtered to 1.78 Hz, and then downsampled to 15 Hz. We performed the extraction of the respiratory rate on every 1-min segment of PS, and then the data were shifted by every 10 s for the entire 3 min of recordings, i.e., each 1-min dataset had a 50 s overlap. Thus, for each 3-min segment, we had thirteen 1-min segments to analyze for all methods to be compared. Thus, 3 min of data were sufficiently long to test the efficacy of each method but not too long in duration to fatigue the subjects as their task was to breathe on cue with a metronome-timed beep sound. For the VFCDM and CWT methods, for every 1-min segment, the initial and final 5 s of the TFS were not considered because the TFS has an inherent end effect which leads to inaccurate time–frequency estimates. For the CWT method, the lower and upper frequency bounds of the analyzed signal were set to 0.01 and 0.5, respectively. The filter parameters of the VFCDM were set to the first cutoff frequency $F_w = 0.03$ Hz, second cutoff frequency $F_v = 0.015$ Hz, and filter length $N_w = 64$. We have previously shown that the parameter $F_v = F_w/2$, and that N_w is chosen to be approximately half the data length. For each of these categories, detection errors were found for each frequency for all subjects using the four different methods. The error ε is calculated as follows:

$$\varepsilon = \frac{\sum_{i=1}^{n} |R_D^i - R_T^i|}{n}, \tag{7}$$

where n is the number of 1-min segments, R_D^i and R_T^i denote the detected breathing rate and the true breathing rate of i-th 1-min dataset, respectively.

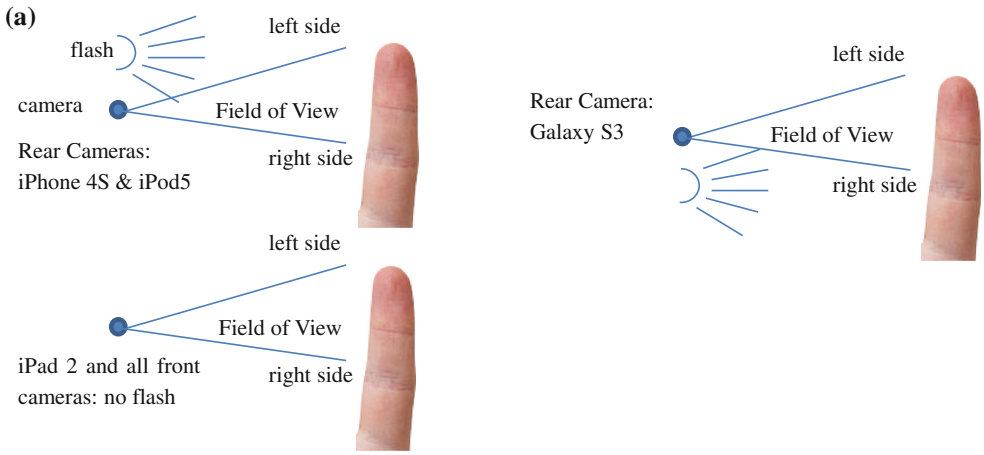
RESULTS

Selection of the Best Color Band and the Optimal Video Pixel Size for Estimation of Heart Rates

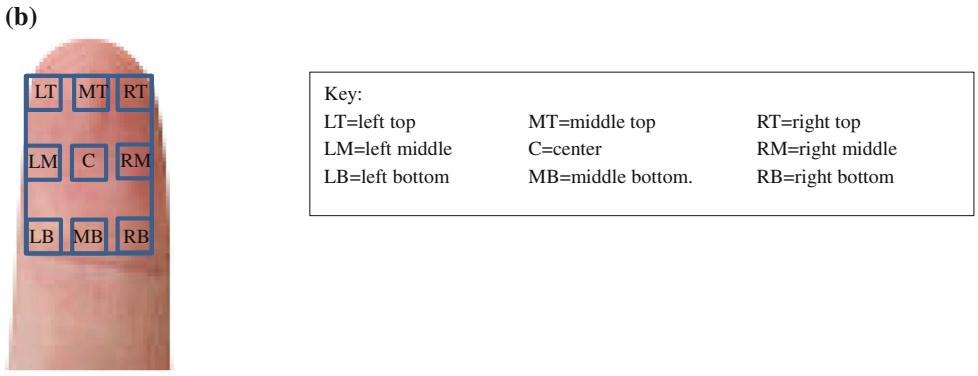
Figure 2a shows the orientation of the Field of View (FOV) of each camera relative to the location of the camera flash. All references to "left" and "right" in this paper are relative to the camera FOV, regardless of whether the camera itself was on the front or rear of the device. Note that when a device's front video camera is on, what is displayed in the LCD display of the device is a mirror image of the FOV of the front camera. The stored video will revert to the FOV view, but until the videotaping is complete, the display in the LCD of the device will be the mirror image of the actual front camera FOV. This is to match people's expectations as they look in the display while photographing themselves. However, reversal in the display was not taken into account to avoid confusion, and because we used the video feed directly before it was processed for display on the device's LCD.

Figures 2b and 2c provide details of the video pixel regions examined on all four devices and they consist of the following 12 video regions: left top (LT), left middle (LM), left bottom (LB), right top (RT), right middle (RM), right bottom (RB), middle top (MT), center (C), middle bottom (MB), vertical left half-VGA (vertical left HVGA), vertical right half-VGA (vertical right HVGA) and VGA.

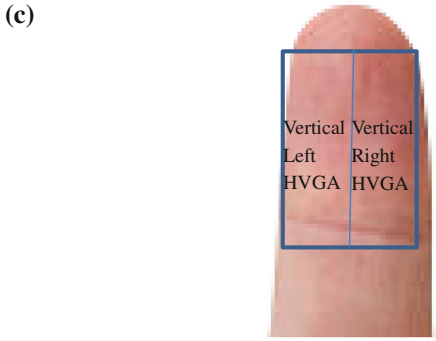
All results shown are based on average values from 10 subjects. When the flashlight was on (back camera



Position of Flash Relative to Camera Field of View



Selected 50x50 Pixel Regions (not to scale) within Camera's Field of View – All Devices Except iPad, which had Right on Top in Landscape



Division of Camera's Field of View into Vertical Left Half VGA & Right Half VGA

FIGURE 2. Example of different regions of iPhone 4S, iPad 2, iPod 5, and Galaxy S3. The top panel (Fig. 2a) represents the camera's FOV and relative position of flash LED's. The middle panel (Fig. 2b) shows the locations of the 9 50 × 50 pixel regions in the camera's FOV. The bottom panel (Fig. 2c) shows the division of the FOV into left and right vertical halves, each of HVGA resolution.

displays for iPhone 4S, iPod 5 and Galaxy S3), the green color consistently provided significantly higher mean amplitude values than either the blue or red color.[19] Table 1 shows experimental results of R–R intervals (RRIs) extracted from ECG and three-color band PS from an iPhone 4S. As shown in Table 1, the PS values from the smart phone are an excellent surrogate to RRIs derived from ECG for all colors. There was no statistical difference between RRIs derived from ECG and each of the three color PS; the median

TABLE 1. Experimental results of heart rate extracted from ECG and three-color band signals obtained from iPhone 4S (Resolution: HVGA).

Color	PS	RRI	Median error
Blue	0.8124 ± 0.23334	0.8103 ± 0.0514	0.0021
Green	0.8149 ± 0.19698		0.0047
Red	0.8121 ± 0.22897		0.0018

errors calculated using Eq. (1) are also very small for all three color band signals. Figure 3 shows the Bland–Altman plot for the mean HR data from the iPhone 4S (green color) and the ECG. The Bland–Altman plot shows a mean difference of 0.074 and that most of the data are within the 95% confidence intervals.

Having established that the green color signal provides the best signal amplitude values for an iPhone 4S, we now systematically investigate which pixel regions of the green color give the most optimal signal quality as determined by the largest amplitude values for all four devices. Specifically, nine different regions of 50 × 50 pixels, the left and right pixel regions of HVGA, and the entire VGA pixel region were investigated for the best signal quality. Table 2 shows the mean amplitude values of the green color pulse signal for different pixel regions of the four devices. For iPhone 4S, the left region of HVGA had the largest amplitude value among the twelve regions, as expected, since the LED flash is placed on the left side of the camera's FOV (see Fig. 2a). For the iPad 2, the device was held vertically on a desk, in landscape mode, so we chose also to consider the FOV in landscape mode. In this case, the right side of the portrait mode FOV was turned to be on top, and the left side was on the bottom. The RT and RM regions of 50 × 50 pixels and the right region of HVGA have among the largest amplitude values since the light source was from the ceiling of the room, i.e. closest to the top in landscape mode. For the iPod 5, the LT and LM regions of 50 × 50 pixels and the VGA have the largest amplitude values. All left values exceed right values because the flash is on the left side of the camera's FOV (see Fig. 2). For the Galaxy S3, the RT, RM and RB regions have the largest amplitude values among the twelve regions as expected since the LED flash is placed to the right of the camera's FOV (see Fig. 2). Hence, depending on the location of the LED flash, the left or right HVGA or 50 × 50 regions of the green color PS have the highest intensity value among all regions tested.

Heart Rate, Frequency Spectrum and Power Spectrum

Figures 4a–4c show an example of a representative 1-min segment of iPhone 4S PS data, its TFS of the green band signal *via* the VFCDM, and the PSD of the AM and FM signals derived from the HR frequency band (e.g., ~1 to 1.5 Hz), respectively, while a subject was breathing at a metronome rate of 18 breaths/min. Note the similarity of the PS in Fig. 4a to those of commercially-available PPG signals. As shown in Fig. 4c, the PSD of the extracted AM and FM time series show the largest peaks at 0.3 Hz; these peaks correspond accurately to the true respiratory rate of 18 breaths/min.

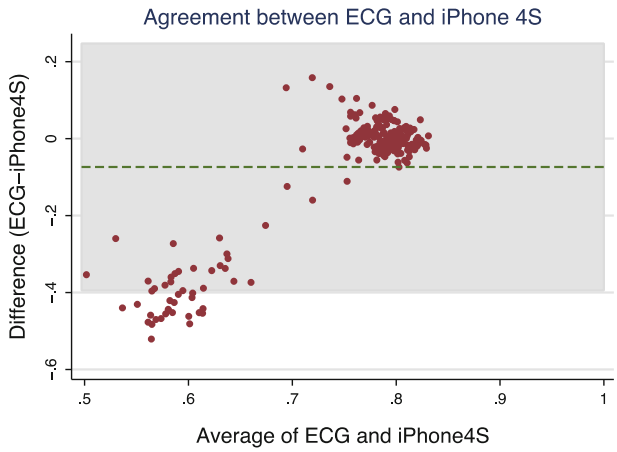


FIGURE 3. Example Bland–Altman plot with a mean difference of 0.074 that shows the limit of agreement of 95% (dashed line is the mean difference ± the limit of agreement) between the continuous HR of a smart phone and the patient's corresponding ECG signal.

Respiratory Rate

The true respiratory rates were derived by taking the PSD of the respiratory impedance trace signals during metronome breathing experiments. True respiratory rates from the respiration trace and the estimated breathing rates from the green signal using both the FM and AM sequences from the VFCDM and CWT were compared using metronome rates ranging from 0.2 to 0.9 Hz. In order to evaluate the four computational methods, we provide figures and tables that show the accuracy and repeatability of each method as a function of the true breathing rate. For tabulating results, we grouped the results for 0.2–0.3 Hz together and designated them as the low frequency (LF) breathing rates. Likewise, the results for 0.4–0.6 Hz breathing rates were lumped together and designated as the high frequency (HF) breathing rates. Since the percentage errors were found to be not-normally distributed, we report the median and inter-quarter range (IQR) values.

Figure 5 shows the subjects' variations of percentage detection error in the form of box plots for the left region of the HVGA pixel resolution with flash on

TABLE 2. The mean amplitude values of the green color pulse signals with flash on except for iPad 2.

			Mean amplitude value			
No.	Resolution	Region	iPhone 4S	iPad 2	iPod 5	Galaxy S3
1	50 × 50	RT	6.33 ± 1.99	4.78 ± 1.42*	2.67 ± 0.82	9385.85 ± 3140.96*
2		RM	7.02 ± 2.19	4.77 ± 1.42*	2.41 ± 0.75	9326.86 ± 3123.12*
3		RB	6.15 ± 1.94	2.44 ± 0.72	2.31 ± 0.72	8583.78 ± 2839.43*
4		MT	8.45 ± 2.64	4.10 ± 1.22	4.11 ± 1.27	7066.07 ± 2365.34
5		Center	9.05 ± 2.82	3.88 ± 1.16	2.79 ± 0.88	6550.41 ± 2173.4
6		MB	8.28 ± 2.59	3.07 ± 0.91	3.59 ± 1.12	3459.99 ± 1148.69
7		LT	9.42 ± 2.94	3.53 ± 1.06	5.79 ± 1.79*	5682.13 ± 1910.77
8		LM	10.49 ± 3.26	2.89 ± 0.85	6.23 ± 1.92*	3969.18 ± 1315.59
9		LB	9.61 ± 3.01	4.05 ± 1.21	5.04 ± 1.57	1605.74 ± 525.84
10	HVGA	Right	8.67 ± 2.54	4.74 ± 1.39*	3.53 ± 1.02	7595.58 ± 2521.62
11		Left	11.37 ± 3.32*	3.78 ± 1.11	5.17 ± 1.49	2766.16 ± 915.96
12	VGA	Full	9.05 ± 2.65	3.11 ± 0.91	5.75 ± 1.66*	5168.72 ± 1715.26

*$p < 0.05$ to other pixel regions.

since this region was found to have the best signal quality as shown in Table 2. The top and bottom panels of Fig. 5 represent results for the LF and HF breathing rates, respectively. The lower boundary of the box closest to zero indicates the 25th percentile, a line within the box marks the median, and the upper boundary of the box farthest from zero indicates the 75th percentile. Whiskers (error bars) above and below the box indicate the 90th and 10th percentiles. Therefore, the area of the blue box is an indication of the spread, i.e., the variation in median error (or IQR), across the population. These figures indicate how well the algorithms perform across the entire population. Red crosses represent the 5th and 95th percentiles.

As shown in Fig. 5, the AR model approach is the least accurate followed by CWT-AM, CWT-FM, and VFCDM (both AM & FM approaches) when we consider all breathing frequencies. Note that the variances of the median values as determined by ε [the average respiratory estimation error as defined in Eq. (7)] are significantly lower for both VFCDM and CWT than for AR model approach. Although there was no significant difference in the median error between CWT and VFCDM methods at 0.3 Hz, ε is found to be the lowest for VFCDM- FM at 0.2 Hz. In general, ε is larger for HF than LF breathing rates for all computational methods. For HF breathing rates, ε is lowest for CWT-FM, followed by VFCDM, CWT-AM, and AR model. While there is no significant difference in the variance between VFCDM-FM and CWT-FM, both methods have significantly less variance than either CWT-AM or VFCDM-AM or AR model. Thus, gauging the accuracy as defined by the median errors and their variances, as shown in Fig. 5, we observed that for HF breathing rates, CWT-FM consistently provides significantly lowest median errors and variance values.

As shown in Fig. 5, the subjects' variation of percentage detection errors has been shown in the form of box plots, which were extracted from front cameras of an iPhone 4S and an iPad 2 (no flash), respectively, for the left HVGA region. While not shown, the left HVGA region also had the best signal quality with the flashlight off for an iPhone 4S. The AR model approach is the least accurate followed by CWT and VFCDM methods when we consider all breathing frequencies. For LF breathing rates, there was no significant difference in the median error between VFCDM methods. However, the variances of the median values as determined by ε are significantly lower for both VFCDM and CWT than for AR model approaches. In general, ε is larger in HF than LF breathing rates. For HF breathing rates, ε is lowest for CWT-FM, followed by VFCDM, CWT-AM, and AR model. While there is no significant difference in the variance between VFCDM-FM and VFCDM-AM in LF breathing rate, median errors of VFCDM-FM are significantly lower than that of VFCDM-AM. Thus, gauging the accuracy as defined by the median errors and their variances, as shown in Fig. 5, we observed that for both LF and HF breathing rates, CWT-FM consistently provides the lowest median errors and variance values.

Figure 5 also shows the subjects' variation of percentage detection error in the form of box plots, which were extracted from front cameras of a Galaxy S3 and an iPod 5, respectively, both from the 50 × 50 pixel resolutions in the LT for the former and LM regions for the latter. The AR model approach is the least accurate followed by CWT and VFCDM methods when we consider all breathing frequencies. For LF breathing rates, there was no significant difference in the median error between VFCDM methods. However, the variances of the median values as determined

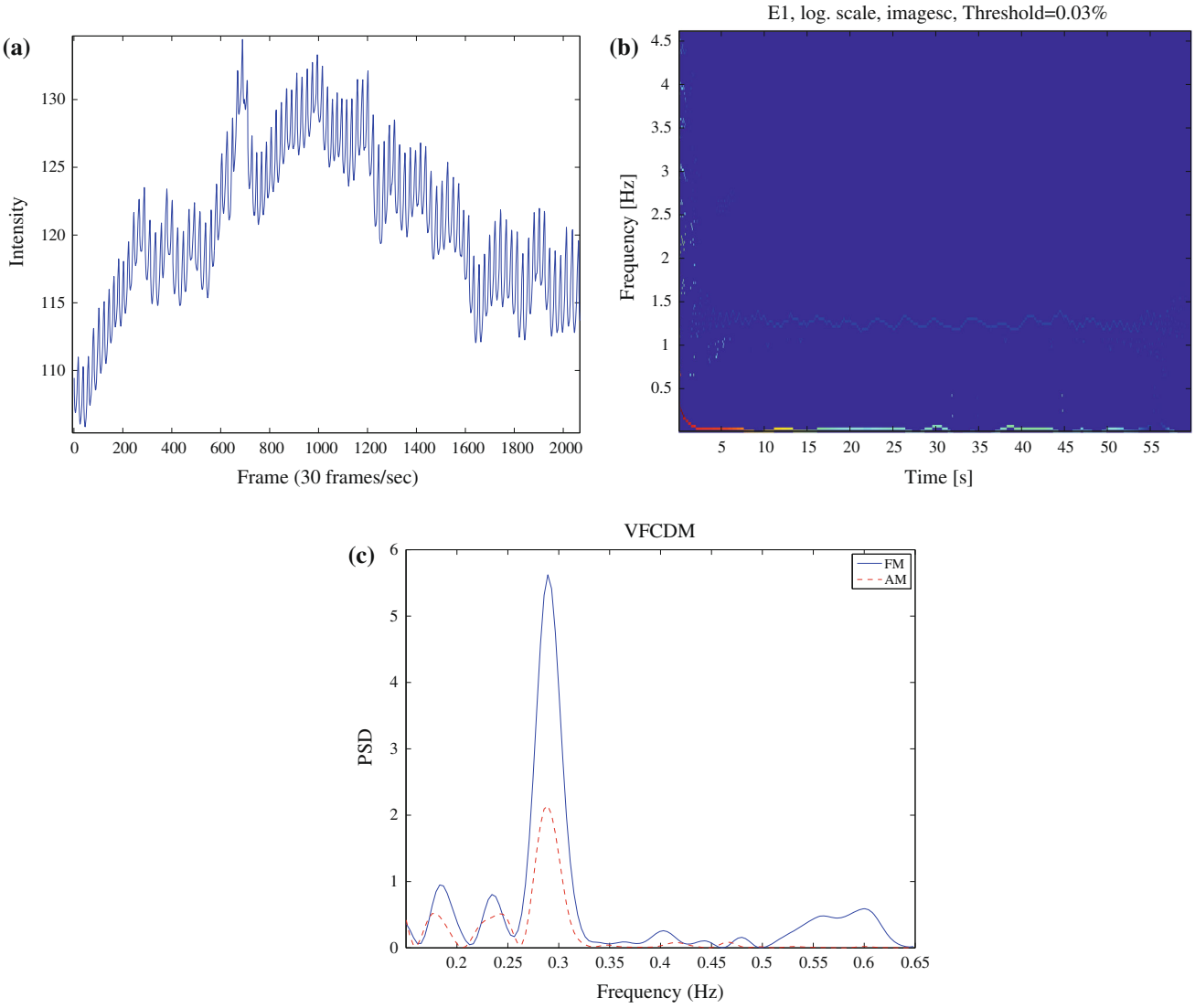


FIGURE 4. PPG signal, estimated instantaneous frequencies, and PSD. (a) Pulsatile signal, (b) Estimated instantaneous frequencies using VFCDM with prominent frequency oscillations seen near heart rate (1.3 Hz), and (c) PSD of PS signal.

by ε are significantly lower for both VFCDM and CWT than for AR model approaches. ε is larger in HF than LF breathing rates. For HF breathing rates, ε is lowest for CWT-FM. While there is no significant difference in the variance between VFCDM-FM and VFCDM-AM in LF breathing rate, median errors of VFCDM-FM are significantly lower than that of VFCDM-AM. Thus, gauging the accuracy as defined by the median errors and their variances, as shown in Fig. 5, we observed that for both LF and HF breathing rates, VFCDM-FM most often provides the lowest median errors and variance values.

Table 3 shows the numerical statistics (IQR) for the "repeatability" across the population of test subjects. The results for 0.2–0.4 Hz (LF breathing range) breathing rates are much better than for 0.5–0.6 Hz (HF breathing range), and in addition, the tracking ability of the breathing rate detection method is much better when CWT and VFCDM methods are used for the LF. Even though the AR method shows significantly lower values of IQR errors than all the other methods studied here, the AR method provided relatively high median errors. For each of the four different devices, the VFCDM-FM method has significantly lower IQR errors ($\varepsilon < 7$) and median errors ($\varepsilon < 6$) than those of any other devices in the 0.2–0.4 Hz breathing rate range.

ANOVA and the Bonferroni t test were used for analysis of differences between the medians for the seven different methods. Statistical significance was identified as $p < 0.05$. Tables 4 and 5 provide a summary of the statistical analysis comparing the performance of the five methods (AR, CWT-AM, CWT-FM, VFCDM-AM and VFCDM-FM) to each other. For

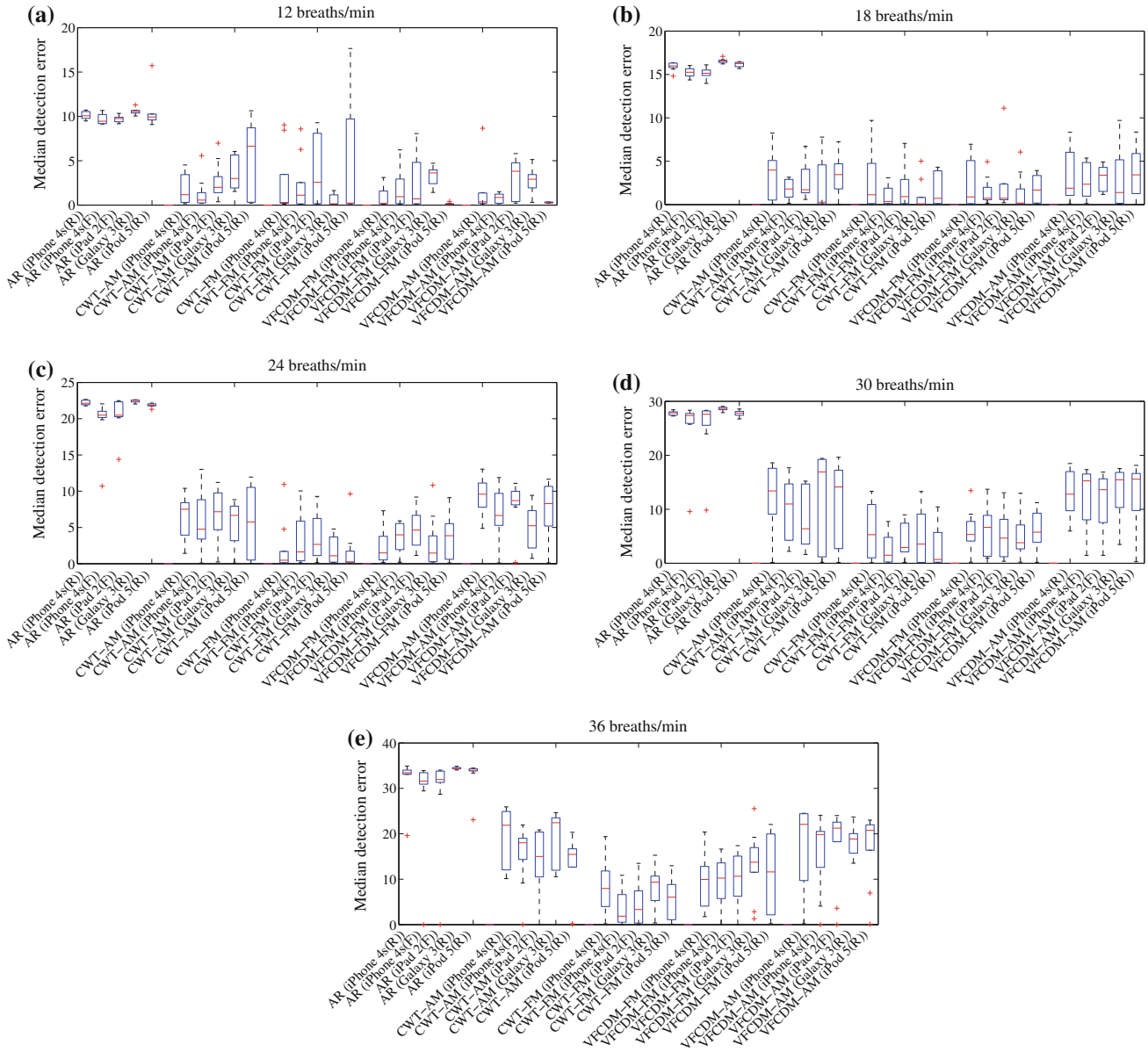


FIGURE 5. Median and IQR errors measured from iPhone 4S, iPad 2, Galaxy S3, iPod 5 when the flashlight was turned on and off. (a)–(e) represent LH (12 and 18 breaths/min) and HF (24, 30 and 36 breaths/min) breathing rates, respectively.

Tables 4 and 5, we list only those comparison that show significant difference among the five computation methods for each device for both LF and HF breathing ranges. Regarding accuracy, for both LF and HF breathing ranges, the tables show that the AR is significantly less accurate than either the AM or the FM version of the CWT and VFCDM methods for all four devices. Further, FM of CWT and VFCDM are significantly more accurate than their AM counterparts for all four devices but only for the HF breathing ranges. The repeatability values shown in Tables 5 are similar to the accuracy results. For example, for both LF and HF breathing ranges, the AR is significantly less repeatable than either AM or FM of CWT and VFCDM methods for all four devices. For the HF breathing range, FM of CWT and VFCDM are significantly more repeatable than their AM counterparts for all four devices.

Table 6 summarizes these measures of median and IQR errors for 0.7, 0.8, and 0.9 Hz breathing rates—rates above what we termed HF rates. As presented numerically in the table, we observe that WT-FM provides the lowest median error at the 0.7 Hz breathing rate, and might be acceptable. However, no method provided reasonably good estimates of breathing rates above the 0.7 Hz breathing rate.

TABLE 3. Population statistics for IQR detection errors for each method.

			WT		CDM	
Device	Breaths/min	AR	AM	FM	FM	AM
iPhone 4S	12	1.06 ± 0.55	1.52 ± 0.8	3.65 ± 1.89	3.17 ± 1.65	1.04 ± 0.53
	18	0.94 ± 0.47	2.25 ± 1.12	2.08 ± 1.14	1.84 ± 1.01	3.95 ± 2
	24	1.28 ± 0.65	6.12 ± 3.12	5.86 ± 3.08	3.76 ± 1.89	5.24 ± 2.67
	30	1.95 ± 1.02	11.54 ± 5.8	4.82 ± 2.5	8.87 ± 4.47	9.03 ± 4.86
	36	2.48 ± 1.32	4.57 ± 2.43	6.38 ± 3.46	7.02 ± 3.51	7.44 ± 3.94
iPad 2	12	0.59 ± 0.3	2.69 ± 1.38	7.96 ± 4.08	5.18 ± 2.84	4.58 ± 2.39
	18	0.83 ± 0.42	3.03 ± 1.63	3.66 ± 1.92	1.89 ± 1.03	2.84 ± 1.45
	24	2.15 ± 1.17	5.94 ± 2.98	6.25 ± 3.22	4.4 ± 2.2	2.01 ± 1.02
	30	3.21 ± 1.7	11.24 ± 5.83	5.98 ± 3.2	8.01 ± 4.01	9.2 ± 4.8
	36	2.45 ± 1.28	8.93 ± 4.48	6.95 ± 3.54	9.15 ± 4.6	4.34 ± 2.23
Galaxy S3	12	0.42 ± 0.22	1.26 ± 0.64	2.1 ± 1.05	1.68 ± 0.92	1.09 ± 0.55
	18	0.41 ± 0.21	4.31 ± 2.31	5.99 ± 3	6.04 ± 3.12	4.25 ± 2.2
	24	0.15 ± 0.08	7.96 ± 4.03	6.48 ± 3.24	5.28 ± 2.66	5.79 ± 2.89
	30	0.42 ± 0.22	7.02 ± 3.51	7.55 ± 3.8	5.01 ± 2.58	2.93 ± 1.48
	36	0.69 ± 0.35	9.94 ± 4.97	14.07 ± 7.13	7.79 ± 3.99	7.93 ± 4.18
iPod 5	12	8.64 ± 4.72	7.48 ± 3.79	4.42 ± 2.21	3.29 ± 1.66	5.49 ± 3.05
	18	0.4 ± 0.2	3.88 ± 1.97	4.96 ± 2.59	4.09 ± 2.06	2.1 ± 1.07
	24	0.67 ± 0.34	4.54 ± 2.27	2.38 ± 1.29	4.57 ± 2.4	7.87 ± 4.05
	30	0.38 ± 0.19	5.38 ± 2.77	6.57 ± 3.29	6.77 ± 3.46	12.21 ± 6.51
	36	0.9 ± 0.45	9.34 ± 4.68	19.38 ± 9.97	11.19 ± 5.74	11.8 ± 5.96

The error values listed for each method represent breaths/min.

TABLE 4. Statistical significance (accuracy) among the five methods for four devices.

Device	LF	HF	Device	LF	HF
iPhone 4S	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. VFCDM-FM VFCDM-AM vs. CWT-FM CWT-AM vs. VFCDM-FM CWT-AM vs. WT-FM	iPod 5	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM CWT-AM vs. VFCDM-FM	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. VFCDM-FM VFCDM-AM vs. CWT-FM CWT-AM vs. WT-FM
iPad 2		AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. VFCDM-FM VFCDM-AM vs. CWT-FM CWT-AM vs. CWT-FM	Galaxy S3	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. VFCDM-FM VFCDM-AM vs. CWT-FM CWT-AM vs. CWT-FM

Figure 6 shows the subjects' variation of percentage detection error in the form of box plots extracted from a rear camera (with flashlight on) of an iPhone 4S during spontaneous breathing. True respiration rate was found by computing the PSD of the impedance respiration trace signal and finding the frequency at the maximum amplitude using a respiration belt. The variances of the median values as determined by ε are significantly lower for both VFCDM and CWT than for the AR model approach. In the normal range (11–27 breaths/min), VFCDM-FM consistently provides the lowest median errors and variance values. As shown in Table 7, there was no significant difference in the median error among WT-AM, WT-FM, VFCDM-FM, and VFCDM-AM during spontaneous breathing, the accuracy of AR is lower than other approaches.

In general, the ability of the methods to provide consistent results is especially excellent (highest) for both the CWT-FM and VFCDM methods, for both LF and HF breathing rates. As with the accuracy results, the repeatability is also better for the LF than for the HF breathing rates for all four methods. Both CWT-FM and VFCDM provide significantly more repeatable results than either CWT-AM or AR model.

TABLE 5. Statistical significance (repeatability across time) among the five methods for four devices.

Device	LF	HF	Device	LF	HF
iPhone 4S	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. CWT-FM	iPod 5	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM	AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. CWT-FM
iPad 2		AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM VFCDM-AM vs. CWT-FM	Galaxy S3		AR vs. VFCDM-AM AR vs. VFCDM-FM AR vs. CWT-AM AR vs. CWT-FM

TABLE 6. Accuracy as determined by median errors at 42, 48, 54 breaths/min (iPhone 4S, flashlight: On). The error values listed for each method represent breaths/min.

			WT		CDM	
Breaths/min	Error	AR	AM	FM	FM	AM
42 (0.7 Hz)	Median	40.05 ± 0.41	21.58 ± 9.14	5.58 ± 5.16	16.05 ± 4.58	24.21 ± 6.33
	IQR	0.72 ± 0.38	19.89 ± 10.15	7.22 ± 3.87	9.27 ± 4.72	5.17 ± 2.59
48 (0.8 Hz)	Median	45.69 ± 1.21	32.61 ± 4.65	24.06 ± 9.67	24.74 ± 4.08	28.53 ± 6.82
	IQR	0.68 ± 0.35	9.3 ± 4.97	14.07 ± 7.04	4.61 ± 2.32	6.25 ± 3.15
54 (0.9 Hz)	Median	51.49 ± 1.46	38.14 ± 4.9	36.38 ± 3.55	32.8 ± 4.87	33.24 ± 8.93
	IQR	0.41 ± 0.22	6.79 ± 3.68	6.93 ± 3.51	6.05 ± 3.07	11.77 ± 6.28

Computation Time

Table 8 shows the computational time for heart rate extraction based on the choice of pixel resolution and the number of color bands used. As shown in the table, pixel resolutions of QVGA and HVGA result in a frame rate of 25 frames/s when only one color is selected. The frame rates extracted from two and three colors are 23 and 20 frames/s, respectively, in the case of HVGA resolution.

The clock speed of the CPU used in the iPhone 4S and iPod 5 is 800 MHz. The latest iPhone 5 is clocked at 1.02 GHz. The recently released Samsung Galaxy S4 is equipped with a 1.9 GHz Quad-core processor. Thus, for most new smartphone and tablet cameras, higher than 30 frames/s can be achieved, suggesting that a choice of higher pixel resolution will not be a significant problem for accurate and real-time detection of heart rates and respiratory rates.

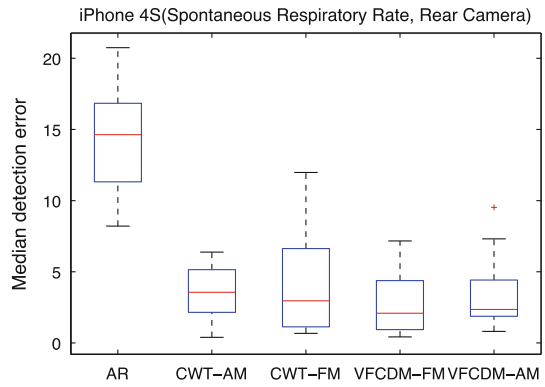


FIGURE 6. Spontaneous respiratory rate.

TABLE 7. Statistical significance (accuracy and repeatability across time) among the five methods for spontaneous respiratory rate.

Accuracy	Repeatability across Time
AR vs. VFCDM-AM	AR vs. VFCDM-AM
AR vs. VFCDM-FM	AR vs. VFCDM-FM
AR vs. WT-AM	AR vs. WT-AM
AR vs. WT-FM	AR vs. WT-FM

DISCUSSION

In this work, we tested several smartphones and tablets for their feasibility in estimating respiratory rates using the PS derived from a resident video camera and flashlight, when available. The motivation for this work is based on several recent works which showed that accurate respiratory rates, especially at normal breathing rates, can be obtained from pulse oximeters.[11–13] The characteristics of the PS derived from cameras in smartphones and tablets are similar to PPG signals, hence, similarly-accurate respiratory rates can be obtained, theoretically. Our results do indicate

TABLE 8. Computation time of heart rate extracted from color band signal of iPhone 4S depending on different resolutions.

Resolution	Color	Computation time
320 × 240 (QVGA)	Green	25 frames/s
480 × 320 (HVGA)	Green	25 frames/s
480 × 320 (HVGA)	Green and red	23 frames/s
480 × 320 (HVGA)	3 Colors	20 frames/s
640 × 480 (VGA)	Green or red	19 frames/s

that certainly for normal breathing ranges (0.2–0.3 Hz), this is feasible from PS derived from smartphone and tablet video cameras.

We have optimized the accuracy of the respiratory rates by first systematically analyzing the optimal pixel resolution of the video signal for the attainment of the strongest PS strength. It is logical to assume that the greater the amplitude of the PS, the higher the signal's strength with the proviso that care is taken to minimize motion artifacts during measurements. Our results showed that a choice of larger pixel resolutions does not necessary result in higher PS amplitude. For example, for the Galaxy S3, iPod 5 and iPad 2, 50 × 50 resolution provided either the highest pulsatile amplitude or was statistically equivalent to HVGA resolution. In fact, HVGA resolution was the best choice only for the iPhone 4S. The important implication of having a smaller pixel region providing just as good or better signal quality than a larger pixel region is the significant reduction in the computational time so that real-time calculation of respiratory rates can be attained.

Commercial pulse oximeters in either transmittance or reflectance mode normally employ a single photodetector (PD) element, typically with an active area of about 6–10 mm^2. The image sensor size of the iPhone 4S is $4.54 \times 3.42 = 15.5268$ mm^2. Consequently, when signals are extracted from HVGA (320 × 480 pixels) video mode, the active area is $2.27 \times 3.42 = 7.7634$ mm^2. Hence, we initially thought that motion artifact and noise can be significantly reduced by increasing the active area in the sensor. However, our investigation revealed that larger pixel resolutions do not necessary result in a higher signal-to-noise ratio.

We compared AR-based approaches, CWT, and VFCDM for respiratory rate estimation from smartphones and a tablet because these techniques have been shown to provide good results from PPG signals. Similar to PPG signal results, the VFCDM-FM provided the most accurate respiratory rate estimation with the fastest computational time than any of the methods compared in this study for the LF breathing rate. For HF breathing rates, both CWT and VFCDM methods provided comparable results. The CWT approach using either the FM or AM signals fared better than the AR method but at the expense of higher computational time.

Due to the inherent non-stationarity in the respiratory rate, a time–frequency method is needed and appears to be the most appropriate approach. Another advantage of the TFS approach to estimating respiratory rates is that unlike most filtering approaches, tuning of a number of parameters specific to each subject is not required. Note that in our work, we have used the same parameters (as described in Methods section) for both CWT and VFCDM for all subjects and for all breathing rates.

As was the case with respiratory rate estimation using the PPG signal, the computational speed of the VFCDM method is faster than that of the wavelet method for smartphone and tablet data. The average time to calculate the respiration frequency using the VFCDM method was found to be around 1.4 s, while using the wavelet method took 37.8 s on average (programs running on MATLAB R2012a). The AR spectral method was the fastest as it took only 0.2 s on average using MATLAB, and this computation time includes the time needed to calculate the model order based on an initial model order selection of 50. However, the AR method is the least accurate in respiratory rate estimation.

All three methods showed increased estimation errors as the breathing rates increased, for all devices tested. This observation was also noted for the PPG signal.[2] We have also examined breathing rates of 0.7 Hz, 0.8 Hz and 0.9 Hz, and the results showed significant deviation from the true breathing rates for all 3 methods. Both CWT and VFCDM methods provided comparable results with significantly worse estimates for the AR method which was also the case with both LF and HF breathing rates. Hence, our results show that it is feasible to obtain good results for the normal breathing rates but not higher breathing rates (i.e., >0.5 Hz). We can speculate that there are two reasons for inaccurate results for high breathing rates. First, detection of both AM and FM phenomenon requires persistent oscillations for several cycles, but with faster respiratory rates, our decision to limit the data segment to 1 min may not be sufficient. More importantly, with faster breathing rates, the AM or FM phenomenon becomes less apparent, and thus, it becomes more difficult to detect them no matter how sophisticated the detection may be.

We have not considered the device-to-device variations in obtaining respiratory rates. However, we do not believe this is a concern because the specifications of the camera from one device to another device is tightly controlled by the phone manufacturers and hence should not vary at all, and if so, it should only be

a minimal amount. Moreover, the pixel resolutions of the examined cameras have more than sufficient resolution to resolve pulse changes, hence, small variations in the pixel resolution, if they occur, should not really affect the respiratory rate estimates. Thus, we believe the device-to-device variation is minimal or not at all, thus, it should not affect the respiratory rate estimation.

In summary, our work was undertaken to determine the optimal pixel resolution and location as well as the color band for obtaining the best quality signal so that we maximize the accuracy of respiratory rate estimation from a video signal from either smartphones or tablets. It was found that a larger pixel resolution does not necessarily result in better signal quality. In fact in most scenarios, a 50 × 50 pixel resolution was just as good as or better than HVGA resolution. In addition, we found that the region closest to the flash in most cases resulted in a higher signal quality which is logical and expected. Finally, using the optimum pixel size, location and color band of the PS, we found accurate respiratory estimates especially in the normal breathing ranges. We found increased breathing rate estimation errors as the respiratory rates increased higher than 0.5 Hz with unreliable results at 0.6 Hz or higher. When both computational time and estimation accuracy are taken into account, the VFCDM-FM provided the best results among all approaches examined in this work. This work allows attainment of at least two vital sign measurements all directly from a finger pressed onto a video camera of either a smartphone or tablet: the heart rate and respiratory rate. It is expected that future work by either our laboratory or others will result in additional other vital sign capabilities directly from a video signal acquired from either a smartphone or tablet.

ACKNOWLEDGMENTS

This work was supported in part by the US Army Medical Research and Materiel Command (USAMRMC) under Grant No. W81XWH-12-1-0541.

REFERENCES

[1]Allison, R., E. Holmes, and J. Nyboer. Volumetric dynamics of respiration as measured by electrical impedance plethysmography. *J. Appl. Physiol.* 19:166–173, 1964.

[2]Chon, K. H., S. Dash, and K. Ju. Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation. *Biomed. Eng. IEEE Trans.* 56:2054–2063, 2009.

[3]Fieselmann, J. F., M. S. Hendryx, C. M. Helms, and D. S. Wakefield. Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. *J. Gen. Intern. Med.* 8:354–360, 1993.

[4]Grimaldi, D., Y. Kurylyak, F. Lamonaca, and A. Nastro. Photoplethysmography detection by smartphone's videocamera. In: The 6th IEEE International Conference on Intelligent Data Acquisition and Advance Computing Systems: Technology and Applications, 2011.

[5]Hasselgren, M., M. Arne, A. Lindahl, S. Janson, and B. Lundbäck. Estimated prevalences of respiratory symptoms, asthma and chronic obstructive pulmonary disease related to detection rate in primary health care. *Scand. J. Prim. Health Care* 19:54–57, 2001.

[6]Hirsch, J., and B. Bishop. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *Am. J. Physiol. Heart Circ. Physiol.* 241:H620–H629, 1981.

[7]Lee, J., and K. Chon. Respiratory rate extraction via an autoregressive model using the optimal parameter search criterion. *Ann. Biomed. Eng.* 38:3218–3225, 2010.

[8]Lee, J., and K. H. Chon. An autoregressive model-based particle filtering algorithms for extraction of respiratory rates as high as 90 breaths per minute from pulse oximeter. *Biomed. Eng. IEEE Trans.* 57:2158–2167, 2010.

[9]Lee, J., and K. H. Chon. Time-varying autoregressive model-based multiple modes particle filtering algorithm for respiratory rate extraction from pulse oximeter. *Biomed. Eng. IEEE Trans.* 58:790–794, 2011.

[10]Lee, J., J. P. Florian, and K. H. Chon. Respiratory rate extraction from pulse oximeter and electrocardiographic recordings. *Physiol. Meas.* 32:1763, 2011.

[11]Leonard, P., T. Beattie, P. Addison, and J. Watson. Standard pulse oximeters can be used to monitor respiratory rate. *Emerg. Med. J.* 20:524–525, 2003.

[12]Leonard, P. A., D. Clifton, P. S. Addison, J. N. Watson, and T. Beattie. An automated algorithm for determining respiratory rate by photoplethysmogram in children. *Acta Paediatr.* 95:1124–1128, 2006.

[13]Leonard, P., N. R. Grubb, P. S. Addison, D. Clifton, and J. N. Watson. An algorithm for the detection of individual breaths from the pulse oximeter waveform. *J. Clin. Monit. Comput.* 18:309–312, 2004.

[14]Liu, H., Y. Wang, and L. Wang. A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012. doi:10.1109/EMBC.2012.6346371.

[15]Lu, S., K. H. Ju, and K. H. Chon. A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry. *Biomed. Eng. IEEE Trans.* 48:1116–1124, 2001.

[16]Mason, K. P., P. E. Burrows, M. M. Dorsey, D. Zurakowski, and B. Krauss. Accuracy of capnography with a 30 foot nasal cannula for monitoring respiratory rate and end-tidal CO2 in children. *J. Clin. Monit. Comput.* 16:259–262, 2000.

[17]McManus, D. D., J. Lee, O. Maitas, N. Esa, R. Pidikiti, A. Carlucci, J. Harrington, E. Mick, and K. H. Chon. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Heart Rhythm* 10:315–319, 2013.

[18]Rantonen, T., J. Jalonen, J. Grönlund, K. Antila, D. Southall, and I. Välimäki. Increased amplitude modulation of continuous respiration precedes sudden infant death syndrome:–Detection by spectral estimation of respirogram. *Early Hum. Dev.* 53:53–63, 1998.

[19]Scully, C. G., J. Lee, J. Meyer, A. M. Gorbach, D. Granquist-Fraser, Y. Mendelson, and K. H. Chon. Physiological parameter monitoring from optical recordings with a mobile phone. *Biomed. Eng. IEEE Trans.* 59:303–306, 2012.
[20]South, M. Measurement of respiratory rate and timing using a nasal thermocouple. *J. Clin. Monit.* 11:159–164, 1995.
[21]Subbe, C., R. Davies, E. Williams, P. Rutherford, and L. Gemmell. Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions*. *Anaesthesia* 58:797–802, 2003.
[22]Torrence, C., and G. P. Compo. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79:61–78, 1998.
[23]Wang, H., K. Siu, K. Ju, and K. H. Chon. A high resolution approach to estimating time-frequency spectra and their amplitudes. *Ann. Biomed. Eng.* 34:326–338, 2006.
[24]Younes, M. Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders. *J. Appl. Physiol.* 105:1389–1405, 2008.

Annals of Biomedical Engineering (© 2014)
DOI: 10.1007/s10439-014-1030-8

BMES BIOMEDICAL ENGINEERING SOCIETY™ www.bmes.org

Photoplethysmograph Signal Reconstruction based on a Novel Motion Artifact Detection-Reduction Approach. Part II: Motion and Noise Artifact Removal

S. M. A. Salehizadeh,[1] Duy K. Dao,[1] Jo Woon Chong,[1] David McManus,[2] Chad Darling,[3] Yitzhak Mendelson,[1] and Ki H. Chon[1]

[1]Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA; [2]Cardiology Division, Departments of Medicine and Quantitative Health Sciences, University of Massachusetts Medical Center, Worcester, MA 01655, USA; and [3]Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA 01655, USA

(*Received 28 February 2014; accepted 6 May 2014*)

Associate Editor Tingrui Pan oversaw the review of this article.

Abstract—We introduce a new method to reconstruct motion and noise artifact (MNA) contaminated photoplethysmogram (PPG) data. A method to detect MNA corrupted data is provided in a companion paper. Our reconstruction algorithm is based on an iterative motion artifact removal (IMAR) approach, which utilizes the singular spectral analysis algorithm to remove MNA artifacts so that the most accurate estimates of uncorrupted heart rates (HRs) and arterial oxygen saturation (SpO_2) values recorded by a pulse oximeter can be derived. Using both computer simulations and three different experimental data sets, we show that the proposed IMAR approach can reliably reconstruct MNA corrupted data segments, as the estimated HR and SpO_2 values do not significantly deviate from the uncorrupted reference measurements. Comparison of the accuracy of reconstruction of the MNA corrupted data segments between our IMAR approach and the time-domain independent component analysis (TD-ICA) is made for all data sets as the latter method has been shown to provide good performance. For simulated data, there were no significant differences in the reconstructed HR and SpO_2 values starting from 10 dB down to −15 dB for both white and colored noise contaminated PPG data using IMAR; for TD-ICA, significant differences were observed starting at 10 dB. Two experimental PPG data sets were created with contrived MNA by having subjects perform random forehead and rapid side-to-side finger movements show that; the performance of the IMAR approach on these data sets was quite accurate as non-significant differences in the reconstructed HR and SpO_2 were found compared to non-contaminated reference values, in most subjects. In comparison, the accuracy of the TD-ICA was poor as there were significant differences in reconstructed HR and SpO_2 values in most subjects. For non-contrived MNA corrupted PPG data, which were collected with subjects performing walking and stair climbing tasks, the IMAR significantly outperformed TD-ICA as the former method provided HR and SpO_2 values that were non-significantly different than MNA free reference values.

Keywords—Motion artifact removal, Blind source separation, Singular spectrum analysis.

INTRODUCTION

Arterial oxygen saturation reflects the relative amount of oxyhemoglobin in the blood. The most common method to measure it is based on pulse oximetry, whereby oxidized hemoglobin and reduced hemoglobin have significantly different optical spectra. Specifically, at a wavelength of about 660 nm, and a second wavelength between 805 and 960, there is a large difference in light absorbance between reduced and oxidized hemoglobin. A measurement of the percent oxygen saturation of blood is defined as the ratio of oxyhemoglobin to the total concentration of hemoglobin present in the blood. Pulse oximetry assumes that the attenuation of light is due to both the blood and bloodless tissue. Fluctuations of the PPG signal are caused by changes in arterial blood volume associated with each heartbeat, where the magnitude of the fluctuations depends on the amount of blood rushing into the peripheral vascular bed, the optical absorption of the blood, skin, and tissue, and the wavelength used to illuminate the blood.

Address correspondence to Ki H. Chon, Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA. Electronic mail: ssalehizadeh@wpi.edu, dkdao@wpi.edu, jchong@wpi.edu, mcmanusd@ummhc.org, Chad.Darling@umassmed.edu, ym@wpi.edu, kichon@wpi.edu

Published online: 14 May 2014

The pulse oximeter signal contains not only the blood oxygen saturation and heart rate (HR) data, but also other vital physiological information. The fluctuations of photoplethysmogram (PPG) signals contain the influences of arterial, venous, autonomic and respiratory systems on the peripheral circulation. In the current environment where health care costs are ever increasing, a single sensor that has multiple functions is very attractive from a financial perspective. Moreover, utilizing a pulse oximeter as a multi-purpose vital sign monitor has clinical appeal, since it is familiar to the clinician and comfortable for the patient. Knowledge of respiratory rate[3] and HR patterns would be provide more useful clinical information in many situations in which pulse oximeter is the sole monitor available.

Although there are many promising and attractive features of using pulse oximeters for vital sign monitoring, currently they are used on stationary patients. This is mainly because motion and noise artifacts (MNAs) result in unreliable HR and SpO_2 estimation. Clinicians have cited motion artifacts in pulse oximetry as the most common cause of false alarms, loss of signal, and inaccurate readings.[14] A smart watch with a PPG sensor is currently commercially available for monitoring HRs (www.mioglobal.com). However, MNA is a big source of problem for accurate vital extraction from PPG signals and prevents wide adoption of this potentially useful technology for mobile health.

In practice MNA are difficult to remove because they do not have a predefined narrow frequency band and their spectrum often overlaps that of the desired signal.[27] Consequently, development of algorithms capable of reconstructing the corrupted signal and removing artifacts is challenging.

There are a number of general techniques used for artifact detection and removal. One of the methods used to remove motion artifacts is adaptive filtering.[1,5,15,19,24] An adaptive filter is easy to implement and it also can be used in real-time applications, though the requirement of additional sensors to provide reference inputs is the major drawback of such methods.

There are many MNA reduction techniques based on the concept of blind source separation (BSS). BSS is attractive and has garnered significant interest since this approach does not require a reference signal. The aim of the BSS is to estimate a set of uncorrupted signals from a set of mixed signals which is assumed to contain both the clean and MNA sources.[2] Some of the popular BSS techniques are independent component analysis (ICA),[4] canonical correlation analysis (CCA),[28] principle component analysis (PCA),[13] and singular spectrum analysis (SSA).[6]

In ICA, the recorded signals are decomposed into their independent components or sources.[4] CCA uses the second order statistics to generate components derived from their uncorrelated nature.[8] PCA is another noise reduction technique which aims to separate the clean signal dynamics from the MNA data. A multi-scale PCA has been also proposed to account for time-varying dynamics of the signal and motion artifacts from PPG recordings.[20]

A promising approach that can be applied to signal reconstruction is the SSA. The SSA is a model-free BSS technique, which decomposes the data into a number of components which may include trends, oscillatory components, and noise (see historical reviews in Ref. 10). The main advantage of SSA over ICA is that SSA does not require user input to choose the appropriate components for reconstruction and MNA removal. Comparing PCA to SSA, SSA can be applied in cases where the number of signal components is more than the rank of the PCA covariance matrix. Applications of the SSA include extraction of the amplitude and low frequency artifacts from single channel EEG recordings,[26] and removing heart sound dynamics from respiratory signals.[9]

In this paper, we introduce a novel approach to reconstruct a PPG signal from those portions of data that have been identified to be corrupted using the algorithm detailed in Part I of the companion paper. The fidelity of the reconstructed signal was determined by comparing the estimated SpO_2 and HR to reference values. In addition, we compare the reconstructed SpO_2 and HR values obtained *via* the time-domain ICA (TD-ICA) to our method. We have chosen to compare our method to TD-ICA since the latter has recently been shown to provide good reconstruction of corrupted PPG signals.[18]

MATERIALS AND METHOD

Experimental Protocol and Preprocessing

Three sets of data were collected from healthy subjects recruited from the student community of Worcester Polytechnic Institute (WPI). This study was approved by WPI's institutional review board and all the subjects gave informed consent before data recording.

In the first experiment, 11 healthy volunteers were asked to wear a forehead reflectance pulse oximeter developed in our lab along with a reference Masimo Radical (Masimo SET®) finger transmittance pulse oximeter. PPG signals from the forehead sensor and reference (HR) derived from a finger pulse oximeter were acquired simultaneously. The HR and SpO_2 signals were acquired at 80 and 1 Hz, respectively. After baseline recording for 5 min without any movement (i.e., clean data), motion artifacts were induced in the PPG data by the spontaneous movements in both

horizontal and vertical directions of the subject's head while the right middle finger was kept stationary. Subjects were directed to introduce the motions for specific time intervals that determined the percentage of noise within each 1 min segment, varying from 10 to 50%. For example, if a subject was instructed to make left–right movements for 6 s, an 1 min segment of data would contain 10% noise. Note that noise amplitudes varied among subjects due to their head movements.

The second dataset consisted of finger-PPG signals from the nine healthy volunteers in an upright sitting posture using an infrared reflection type PPG transducer (TSD200). An MP1000 pulse oximeter (BIOPAC Systems Inc., CA, USA) was also used to acquire finger PPG signals at 100 Hz. One pulse oximeter of each model was placed on the same hand's index finger (one model) and middle finger (the other model) simultaneously. After baseline recording for 5 min without any movement (i.e., clean data), motion artifacts were induced in the PPG data by the left–right movements of the index finger while the middle finger was kept stationary to provide a reference. This was accomplished by placing the middle finger on top of the edge of a table while the index finger moved in the free space off the edge of the table. Similar to the first dataset, motion was induced at specific time intervals corresponding to 10–50% corruption duration in 1 min segments, i.e., the controlled movement was carried out five times per subject.

The third dataset consisted of data measurements from nine subjects with the PPG signal recorded from the subjects' foreheads using our custom sensor (80 Hz) simultaneously with the reference ECG from a Holter Monitor at 180 Hz and reference HR and SpO_2 derived from a Masimo (Rad-57) pulse oximeter at 0.5 Hz, respectively. The reference pulse oximeter provided HR and SpO_2 measured from the subject's right index finger, which was held steadily to their chest. The signals were recorded while the subjects were going through sets of walking and climbing up and down flights of stairs for approximately 45 min.

Once data were acquired, PPG signals from all three experiments outlined above were preprocessed offline using Matlab (MathWorks, R2012a). The PPG signals were filtered using a zero-phase forward-reverse 4th order IIR band-pass filter with cutoff frequency 0.5-12 Hz.

MOTION ARTIFACT REMOVAL

To reconstruct the artifact-corrupted portion of the PPG signal that has been detected using the support vector machine approach provided in the accompanying paper, we propose a novel hybrid procedure using Iterative Singular Spectrum Analysis (ISSA) and a frequency matching algorithm. Henceforth, we will call these combined procedures the iterative motion artifact removal (IMAR) algorithm.

SINGULAR SPECTRUM ANALYSIS

The SSA is composed of two stages: singular decomposition and spectral reconstruction. The former is the spectral decomposition or eigen-decomposition of the data matrix whereas the latter is the reconstruction of the signal based on using only the significant eigenvectors and associated eigenvalues. The assumption is that given a relatively high signal-to-noise ratio of data, significant eigenvectors and associated eigenvalues represent the signal dynamics and less significant values represent the MNA components.

The calculation of the singular stage of the SSA consists of two steps: embedding followed by singular value decomposition (SVD). In essence, these procedures decompose the data into signal dynamics consisting of trends, oscillatory components, and MNA. The spectral stage of the SSA algorithm also consists of two steps: grouping and diagonal averaging. These two procedures are used to reconstruct the signal dynamics but without the MNA components. In the following section, we detail all four steps in the SSA algorithm.

Singular Decomposition

Embedding

Assume we have a nonzero real-value time series of length N samples, i.e., $X = \{x_1, x_2, \ldots, x_N\}$. In the embedding step, window length $f_s/f_l < L < N/2$ is chosen to embed the initial time series, where f_s is the sampling frequency and f_l is the lowest frequency in the signal. We map the time series X into the L lagged vectors, $X = \{x_i, x_{i+1}, \ldots, x_{i+L-1}\}$ for $i = 1, \ldots, K$, where $K = N - L + 1$.[10] The result is the trajectory data matrix T_x or vector X_i that is each row of T_x for $i = 1, \ldots, K$.

$$T_x = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_K & x_{K+1} & \cdots & x_N \end{bmatrix} \quad (1)$$

From Eq. (1), it is evident that the trajectory matrix, T_x, is a Hankel matrix.

Singular Value Decomposition

The next step is to apply the SVD to the trajectory matrix T_x which results in eigenvalues and eigenvectors

of the matrix $T_x T_x^T$ where T_i for $i = 1, \ldots, L$ can be defined as $T = USV^T$.[10] U_i for $1 < i < L$ is a $K \times L$ orthonormal matrix. S_i for $1 < i < L$ is a diagonal matrix and V_i for $1 < i < L$ is an $L \times L$ square orthonormal matrix, which is considered the principle component. In this step, T_x has L many singular values which are $\sqrt{\lambda_1} > \sqrt{\lambda_2} > , \ldots, \sqrt{\lambda_L}$. By removing components with eigenvalues equal to zero, the ith eigentriple of T_i can be written as $U_i \times \sqrt{\lambda_i} \times V_i^T$ for $i = 1, 2, \ldots, d$, in which $d = \max(i : \sqrt{\lambda_i} > 0)$ is the number of nonzero singular values of T_x. Normally every harmonic component with a different frequency produces two eigentriples with similar singular values. So the trajectory matrix T_x can be denoted as[10]

$$\begin{aligned} T_x &= T_1 + T_2 + \cdots + T_d \\ &= U_1\sqrt{\lambda_1}V_1^T + \cdots + U_d\sqrt{\lambda_d}V_d^T \\ &= \sum_{i=1}^{d} U_i\sqrt{\lambda_i}V_i^T \end{aligned} \tag{2}$$

Projecting the time series onto the direction of each eigenvector yields the corresponding temporal principal component (PC).[7]

Spectral Reconstruction

The reconstruction stage has two steps: grouping and diagonal averaging. First the subgroups of the decomposed trajectory matrices are grouped and then a diagonal averaging step is needed so that a new time series can be formed.[6]

Grouping

The grouping step of the reconstruction stage decomposes the $L \times K$ matrix T_i $(i = 1, 2, \ldots, d)$ into subgroups according to the trend, oscillatory components, and MNA dynamics. The grouping step divides the set of indices $\{1, 2, \ldots, d\}$ into a collection of m disjoint subsets of $I = \{I_1, \ldots, I_m\}$.[12] Thus, T_I corresponds to the group $I = \{I_1, \ldots, I_m\}$. T_{I_i} is a sum of T_j, where $j \in I_i$. So T_x can be expanded as

$$T_x = \overbrace{T_1 + \cdots + T_L}^{SVD} = \overbrace{T_{I_1} + \cdots + T_{I_m}}^{\text{Grouping}} \tag{3}$$

Diagonal Averaging

In the final step of analysis, each resultant matrix, T_{I_i} , in Eq. (3) is transformed into a time series of length N. We obtain the new Hankel matrices $\tilde{X}^{(i)}$ by averaging the diagonal elements of the matrix T_{I_i}.[7] Let H be denoted as the Hankel operator. So that we obtain the Hankel matrix $\tilde{X}^{(i)} = HT_{I_i}$ for $i = 1, \ldots, m$.[12] Under the assumption of weak separability and applying the Hankel procedure to all matrix components of Eq. (3), we obtain the following expansion

$$X = \tilde{X}^{(1)} + \tilde{X}^{(2)} + \cdots + \tilde{X}^{(M)} \tag{4}$$

We can assert that $\tilde{X}^{(1)}$ is related to the trend of the signal; however, harmonic and noisy components do not necessarily follow the order of $\sqrt{\lambda_1} > \sqrt{\lambda_2} > \cdots > \sqrt{\lambda_M}$.

ITERATIVE MOTION ARTIFACT REMOVAL BASED ON SSA

In order to reconstruct the MNA corrupted segment of the signal, an IMAR approach based on SSA was explained in the last section. The ultimate goodness of the reconstructed signal is determined by the accuracy of the estimated SpO_2 and HR values. The top and bottom panels of Fig. 1 show clean and MNA corrupted signals, respectively. The most important part of the SSA is to choose the proper eigenvector components for reconstruction of the signal. Under the assumption of high SNR, the normal practice is to select only the largest eigenvalues and associated eigenvectors for signal reconstruction. However, most often it is difficult to determine the demarcation of the significant from non-significant eigenvalues. Further, the MNA dynamics can overlap with the signal dynamics, hence, choosing the largest eigenvalues does not necessarily result in an MNA-free signal.

To overcome the above limitations, we have modified the SSA approach. The first step of our modified SSA involves computing SVD on both a corrupted data segment and its most prior adjacent clean data segment. Under the assumption of a high SNR of the data, the second step is to retain only the top 5% of the eigenvalues and their associated eigenvectors. The third step is to replace the corrupted segment's top 5% eigenvalues with the clean segment's eigenvalues. The fourth step is to further limit the number of eigenvectors by choosing only those eigenvectors that have HRs between 0.66 Hz $< f_s <$ 3 Hz for both the clean and noise corrupted data segments. The two extreme HRs are chosen so that they account for possible scenarios that one may encounter with low and high HRs. With the remaining candidate eigenvectors resulting from step four, we further prune non-significant eigenvectors by performing frequency matching of the noise corrupted eigenvectors to those of the clean data segment's eigenvectors, in the fifth step. Only those eigenvectors' frequencies that match to those of the clean eigenvectors are retained from the pool of eigenvectors remaining from step four. For the remaining eigenvector candidates, we perform iterative

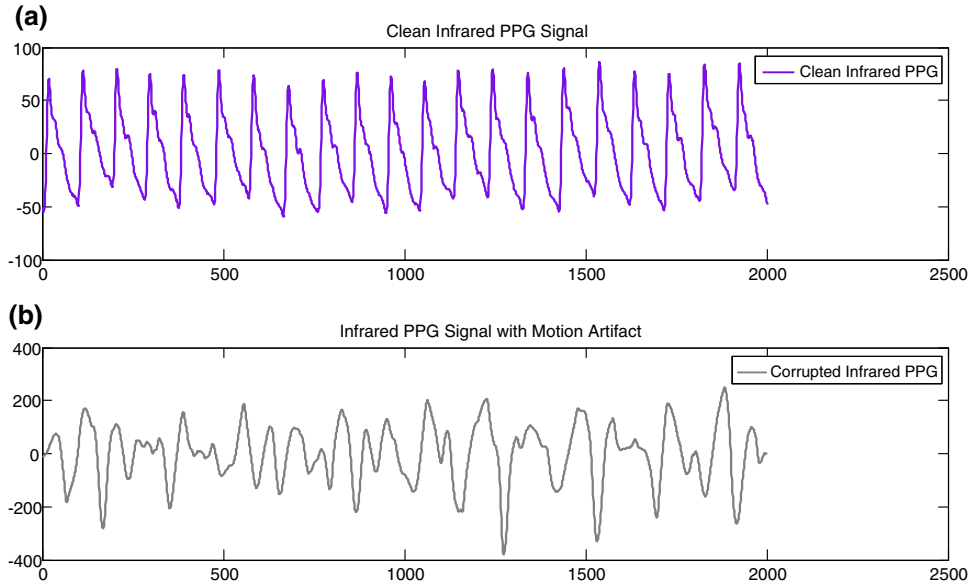


FIGURE 1. Typical infrared PPG signal; (a) clean, (b) corrupted with motion artifacts.

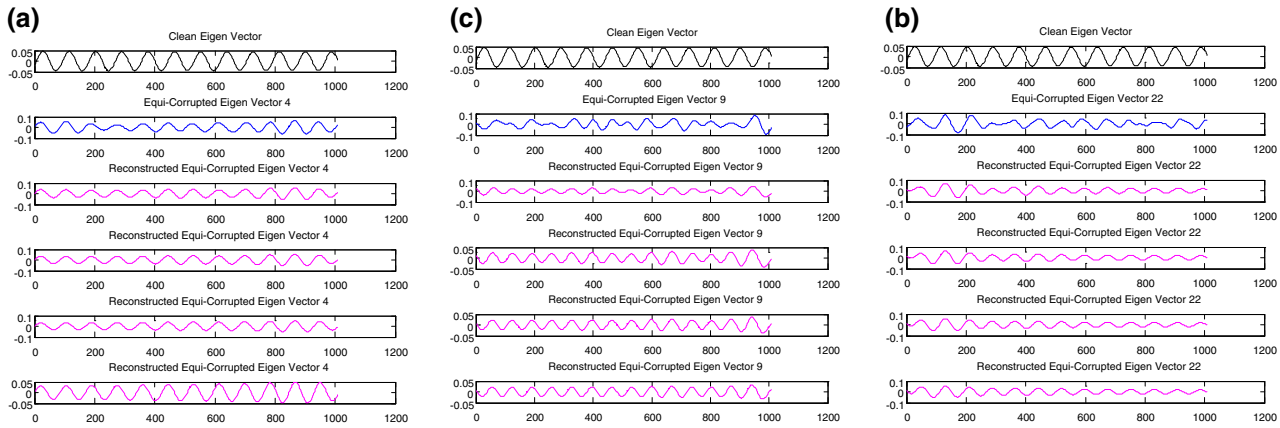


FIGURE 2. Iterative reconstruction of a corrupted eigenvector with frequency of 0.967 Hz. Black font signals (top panels) represent the clean component with frequency of 0.967 Hz; Blue font signals (2nd rows) indicate the corrupted component with the same frequency; Pink font signals are related to iterative evolution of corrupted component to a clean oscillatory signal. (a) Reconstruction of 4th corrupted eigenvector compared to the corresponding clean component. The final pattern after 4 iterations resembles the black font clean component in the top panel. This component is chosen among the components with the same frequency, since it shows the most similarity to the black font clean component. (b) Reconstruction of 9th corrupted eigenvector compared to the corresponding clean component. (c) Reconstruction of 22nd corrupted eigenvector compared to the corresponding clean component.

SSA to further reduce MNA and match the dynamics of the clean data segments' eigenvectors for the final step. For each iteration we perform the standard SSA algorithm. It is our experience that this convergence is achieved within 4 iterations.

Figure 2 shows an example of the iterative SSA procedure applied to candidate eigenvectors that have resulted from step four of the procedure for the modified SSA algorithm. Note that there may be several eigenvectors remaining after the fifth step, hence, this example shows an iterative SSA procedure performed on a particular set of candidate eigenvectors that may match most closely to an eigenvector of a clean data segment. The top row of panels of Fig. 2 represents one of the eigenvectors of the clean signal and the second row of panels represents the MNA corrupted signal's candidate eigenvectors which have the same frequency as that of the clean signal's eigenvector. The remaining lower panels of Fig. 2 represent the candidate eigenvectors after they have gone through four successive iterations of the SSA algorithm. For this portion of the SSA algorithm, we perform SVD on the trajectory matrix of Eq. (1) created from the candidate eigenvector and then reconstruct the eigenvectors

TABLE 1. Iterative motion artifact removal (IMAR) procedure.

Assumption—Heart rate and SpO_2 do not change abruptly and are stationary within the short data segment
Application—Offline Motion Artifact Removal
Objective—Reconstruction of corrupted PPG segment for the purpose of estimating heart rates and SpO_2
Routine
Step 1. First, compute SVD on both corrupted data segments and their most prior adjacent clean data segments
Step 2. Next, keep the top 5% of the clean and corrupted components, based on the eigenvalues being sorted from largest to smallest
Step 3. Replace the corrupted eigenvalues with corresponding clean eigenvalues
Step 4. Among the clean and corrupted components, only choose those with frequency within the heart rate frequency range of $0.66 < F_s < 3$ Hz
Step 5. Apply frequency matching to discard those corrupted components (from Step 4) with different frequencies compared to clean components' frequencies
Step 6. Remove corruption from each component obtained from Step 5 by applying the basic SSA algorithm iteratively
6.a. Calculate the discarding metric for components achieved from SSA iterations and their counterpart clean components from Eq. (5)
6.b. Select those processed components with the closest DM and frequency value to the corresponding clean component's DM and frequency value
Step 7. Finally, reconstruct the corrupted PPG segment based on the components achieved from Step 6

based on SSA using only the first 3 largest eigenvalues obtained from the SVD. This process repeats iteratively until the shape of the reconstructed eigenvector closely resembles one of the clean eigenvectors with the same frequency. It can be seen from Fig. 2 that after 4 iterations the result shown in the (a) panel corresponds most closely to the clean signal's eigenvector, hence, this eigenvector is selected rather than the eigenvectors shown in panels b–c. We calculate the discarding metric (DM) at each iteration and compare this value to the DM value of the corresponding clean component. The DM is calculated according to:

$$DM = \sum |u| \Big/ L(u) \tag{5}$$

where u is the signal component, and $|.|$, $L(.)$ are absolute operator and component length, respectively. The entire procedure for the modified SSA algorithm is summarized in Table 1.

RESULTS

Noise Sensitivity Analysis

To validate the proposed IMAR procedure, we added different SNR levels of Gaussian white noise (GWN) and colored noise to an experimentally collected clean segment of PPG signal. The purpose of the simulation was to quantitatively determine the level of noise that can be tolerated by the algorithm. Seven different SNR levels ranging from 10 dB to −25 dB were considered. For each SNR level, 50 independent realizations of GWN and colored noise were added separately to a clean PPG signal. The Euler–Maruyama method was used to generate colored noise.[17]

Figure 3 shows the results of these simulations with additive GWN. The left panels show pre- and post-reconstruction HR in comparison to the reference HR; the right panels show the corresponding comparison for SpO_2. Tables 2 and 3 show the mean and standard deviation values of the pre- (2nd column) and post-reconstruction (4th column), and the reference (3rd column) HR and SpO_2 values, respectively, for all SNR. The last columns of Tables 2 and 3 also show the estimated HR and SpO_2 values obtained by the TD-ICA method.[18] As shown in Fig. 3 and Tables 2 and 3, the reconstructed HR and SpO_2 values using our IMAR approach were found to be not statistically different when compared to the reference values for all SNR except for −20 and −25 dB. However, the TD-ICA method fails and we obtain significantly different values to those of the reference HR and SpO_2 values when the SNR is lower than −10 dB.

Tables 4 and 5 show corresponding results to that of Tables 2 and 3 but with additive colored noise. Similar to the GWN case, the reconstructed HR and SpO_2 values using the proposed IMAR approach are found to be not significantly different than the reference values for all SNR except for −20 and −25 dB. Moreover, the TD-ICA compares poorly compared to our IMAR as the HR and SpO_2 values from the former method are found to be significantly different to the reference values for all SNR.

Heart Rate and SpO_2 Estimation from Forehead Sensor

As described in "Motion Artifact Removal" section, we collected PPG data under three different experimental settings so that our proposed approach could be more thoroughly tested and validated. For all three experimental settings, the efficacy of our IMAR approach for the reconstruction of the MNA-affected portion of the signal will be compared with the

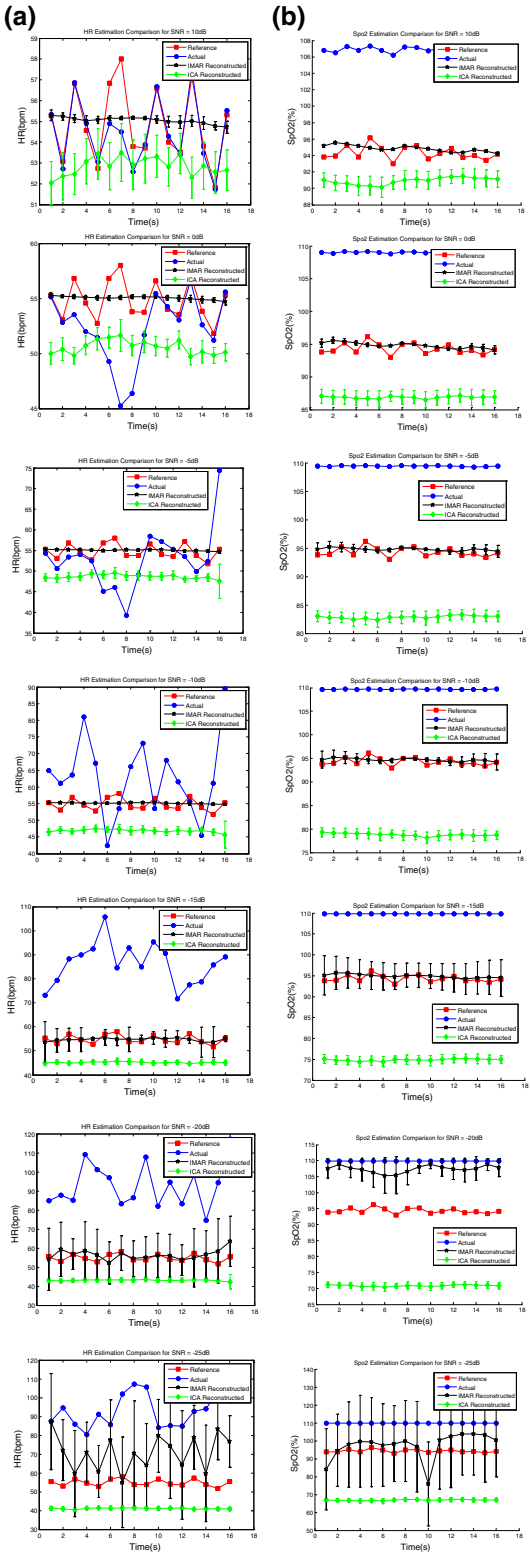


FIGURE 3. (Left) HR estimated from reconstructed PPG for different additive white noise levels; (Right) SpO_2 estimated from reconstructed PPG for different levels of additive white noise.

reference HR and SpO_2 values for all experimental datasets. Unlike Masimo's finger pulse oximeter, our custom-designed forehead pulse oximeter does not calculate moving average HR and SpO_2 values. This is the reason why the standard deviations of HR and SpO_2 values from the forehead sensor are much larger than those from Masimo's finger pulse oximeter (see Tables 6, 7, 8, 9, 10).

For the error-free SpO_2 estimation, Red and IR PPG signals with clearly separable DC and AC components are required. The pulsatile components of the Red and IR PPG signals are denoted as AC_{Red} and DC_{Red}, respectively, and the "ratio-of-ratio" R is estimated[22,23] as

$$R = \frac{AC_{Red}/DC_{Red}}{AC_{IR}/DC_{IR}} \tag{6}$$

Accordingly, SpO_2 is computed by substituting the R value in an empirical linear approximate relation given by

$$SpO_2(\%) = (110 - 25R)(\%) \tag{7}$$

After applying the proposed IMAR procedure to the identified MNA segment of the PPG signal, we estimate the SpO_2 (using Eq. (6, 7)) and HR, and compare it to the corresponding reference and MNA contaminated segment values. As was the case with the "Noise Sensitivity Analysis" section, we compare the performance of the IMAR algorithm to the TD-ICA method. The top and bottom panels of Fig. 4 represent a representative HR and SpO_2 comparison result, respectively. We can see from these figures that the estimated values for both HR (left panels) and SpO_2 (right panels) from the IMAR (black font) track closely to the reference values recorded by the Masimo transmittance type finger pulse oximeter (red square line), while the estimated HR and SpO_2 obtained from the TD-ICA method (green font) deviate significantly from the reference signal. Tables 6 and 7 show comparison of the IMAR and the TD-ICA reconstructed HR and SpO_2 values, respectively, for all 10 subjects. As shown in Table 6, there was no significant difference between the finger reference HR and the IMAR reconstructed HR in 6 out of 10 subjects. However, there was significant difference between the finger reference HR and the TD-ICA reconstructed HR in all 10 subjects. Similarly, the reconstructed SpO_2 values from the IMAR were found to be not significantly different than the finger reference values in 6 out of 10 subjects, but the TD-ICA method was found to be significantly different for all 10 subjects.

TABLE 2. Comparison and statistical analysis of HR estimations from IMAR-reconstructed PPG for different levels of additive white noise.

SNR (dB)	Head HR (mean ± SD)	Finger HR (Reference) (mean ± SD)	IMAR reconstructed HR (mean ± SD)	ICA reconstructed HR (mean ± SD)
10	54.80 ± 2.08	54.81 ± 1.81	55.05 ± 0.15	52.86 ± 0.44*
0	54.80 ± 2.72	54.81 ± 1.81	55.05 ± 0.14	50.58 ± 0.62*
−5	56.37 ± 8.18	54.81 ± 1.81	55.05 ± 0.15	48.64 ± 0.51*
−10	46.02 ± 22.93	54.81 ± 1.81	55.09 ± 0.15	46.85 ± 0.45*
−15	121.62 ± 69.33	54.81 ± 1.81	54.73 ± 0.62	45.17 ± 0.28*
−20	80.08 ± 37.69	54.81 ± 1.81	56.49 ± 2.69	43.08 ± 0.32*
−25	103.62 ± 52.49	54.81 ± 1.81	76.45 ± 7.52*	41.11 ± 0.30*

* $p < 0.05$.

TABLE 3. Comparison & statistical analysis of estimations from IMAR-reconstructed PPG for different levels of additive white noise.

SNR (dB)	Head SpO_2 (mean ± SD)	Finger SpO_2 (Reference) (mean ± SD)	IMAR reconstructed SpO_2 (mean ± SD)	ICA reconstructed SpO_2 (mean ± SD)
10	106.88 ± 0.51	94.23 ± 0.80	94.83 ± 0.38	90.92 ± 0.38*
0	108.98 ± 0.14	94.23 ± 0.80	94.81 ± 0.42	86.88 ± 0.16*
−5	109.42 ± 0.06	94.23 ± 0.80	94.77 ± 0.26	82.86 ± 0.27*
−10	109.69 ± 0.04	94.23 ± 0.80	94.68 ± 0.30	78.81 ± 0.29*
−15	109.82 ± 0.02	94.23 ± 0.80	94.90 ± 0.41	74.88 ± 0.23*
−20	109.89 ± 0.01	94.23 ± 0.80	107.38 ± 1.06*	70.87 ± 0.22*
−25	109.94 ± 0.00	94.23 ± 0.80	97.38 ± 7.39*	66.91 ± 0.26*

* $p < 0.05$.

TABLE 4. Comparison and statistical analysis of HR estimations from IMAR-reconstructed PPG for different levels of additive colored noise.

SNR (dB)	Head HR (mean ± SD)	Finger HR (Reference) (mean ± SD)	IMAR reconstructed HR (mean ± SD)	ICA reconstructed HR (mean ± SD)
10	54.75 ± 1.73	54.81 ± 1.81	55.05 ± 0.26	53.36 ± 0.79
0	55.64 ± 2.72	54.81 ± 1.81	55.06 ± 0.27	50.83 ± 0.54*
−5	55.67 ± 2.88	54.81 ± 1.81	55.06 ± 0.15	48.90 ± 0.32*
−10	51.05 ± 8.24	54.81 ± 1.81	55.07 ± 0.13	46.79 ± 0.30*
−15	61.65 ± 32.08	54.81 ± 1.81	55.17 ± 0.08	45.15 ± 0.30*
−20	73.41 ± 47.73	54.81 ± 1.81	45.96 ± 5.59*	42.96 ± 0.41*
−25	66.37 ± 40.80	54.81 ± 1.81	61.86 ± 2.12*	41.04 ± 0.37*

* $p < 0.05$.

TABLE 5. Comparison and statistical analysis of SpO_2 estimations from IMAR-reconstructed PPG for different levels of additive colored noise.

SNR (dB)	Head SpO_2 (mean ± SD)	Finger SpO_2 (Reference) (mean ± SD)	IMAR reconstructed SpO_2 (mean ± SD)	ICA reconstructed SpO_2 (mean ± SD)
10	94.14 ± 0.99	94.23 ± 0.80	94.85 ± 0.41	90.95 ± 0.18*
0	94.71 ± 1.20	94.23 ± 0.80	94.85 ± 0.53	86.84 ± 0.24*
−5	96.19 ± 1.41	94.23 ± 0.80	93.92 ± 0.83	82.86 ± 0.34*
−10	99.27 ± 1.46	94.23 ± 0.80	94.88 ± 0.96	78.89 ± 0.18*
−15	103.00 ± 0.88	94.23 ± 0.80	94.42 ± 1.71	74.87 ± 0.25*
−20	107.63 ± 0.26	94.23 ± 0.80	74.74 ± 7.92*	70.89 ± 0.17*
−25	105.91 ± 0.49	94.23 ± 0.80	70.75 ± 15.08*	66.89 ± 0.26*

* $p < 0.05$.

TABLE 6. Comparison and statistical analysis of HR estimations from IMAR-reconstructed PPG for 10 different subjects (Head Experiment).

Subject	Head HR (mean ± SD)	Finger HR (Reference) (mean ± SD)	IMAR reconstructed HR (mean ± SD)	ICA reconstructed HR (mean ± SD)
1	68.31 ± 19.25	59.23 ± 1.49	59.76 ± 0.22*	65.68 ± 20.98*
2	85.39 ± 34.53	71.55 ± 3.037	73.72 ± 0.31*	91.02 ± 35.48*
3	76.19 ± 8.88	77.39 ± 1.360	78.705 ± 0.33	68.06 ± 14.14*
4	94.47 ± 39.05	70.55 ± 3.686	73.66 ± 0.38*	75.32 ± 13.42*
5	72.33 ± 29.82	67.88 ± 4.643	66.83 ± 0.39	69.97 ± 20.20*
6	45.09 ± 10.06	51.44 ± 1.481	49.00 ± 0.09*	59.43 ± 22.97*
7	44.82 ± 24.47	59.82 ± 1.486	57.56 ± 0.21	64.49 ± 35.63*
8	63.46 ± 13.35	62.08 ± 0.865	62.23 ± 0.25	60.68 ± 10.70*
9	59.37 ± 30.85	49.05 ± 1.555	49.19 ± 0.20	60.27 ± 13.24*
10	46.89 ± 32.25	79.35 ± 1.323	78.93 ± 0.45	64.80 ± 25.60*

* $p < 0.05$.

TABLE 7. Comparison and statistical analysis of SpO_2 estimations from IMAR-reconstructed PPG for 10 different subjects (Head Experiment).

Subject	Head SpO_2 (mean ± SD)	Finger SpO_2 (Reference) (mean ± SD)	IMAR reconstructed SpO_2 (mean ± SD)	ICA reconstructed SpO_2
1	82.86 ± 4.86	97.70 ± 0.46	97.94 ± 0.93	76.721 ± 38.132*
2	80.33 ± 2.82	97.67 ± 0.47	97.972 ± 4.048*	111.097 ± 1.496*
3	87.20 ± 4.54	95.41 ± 0.49	98.53 ± 0.727*	74.081 ± 21.678*
4	87.36 ± 2.64	97 ± 0	97,13 ± 0.23	81.391 ± 11.81*
5	84.25 ± 3.76	98 ± 0	96.82 ± 5.25*	77.593 ± 22.16*
6	92.38 ± 2.64	98 ± 0	97.47 ± 0.97	84.069 ± 14.84*
7	85.18 ± 3.06	98.41 ± 0.49	96.68 ± 0.38	75.632 ± 17.24*
8	90.94 ± 2.38	99.82 ± 0.06	97.99 ± 0.38	89.322 ± 17.77*
9	83.93 ± 4.54	98 ± 0	99.61 ± 3.87*	100.15 ± 16.96*
10	84.94 ± 4.24	95.97 ± 0.67	96.53 ± 4.62	86.731 ± 19.305*

* $p < 0.05$.

TABLE 8. Comparison and statistical analysis of HR estimations from IMAR-reconstructed PPG for 10 different subjects (Finger Experiment).

Subject	Head HR (mean ± SD)	Finger HR (Reference) (mean ± SD)	IMAR reconstructed HR (mean± SD)	ICA reconstructed HR (mean ± SD)
1	77.43 ± 1.91	70.61 ± 0.73	70.42 ± 0.42	77.32 ± 8.34*
2	63.60 ± 2.42	78.80 ± 0.41	78.36 ± 0.35	79.57 ± 9.68
3	70.82 ± 15.01	66.18 ± 0.76	67.21 ± 0.26	62.96 ± 22.53*
4	87.70 ± 20.53	72.59 ± 0.26	70.85 ± 0.34	73.58 ± 11.34*
5	84.34 ± 4.86	74.43 ± 0.29	73.51 ± 0.29*	77.62 ± 18.55*
6	81.75 ± 6.34	67.78 ± 0.36	69.07 ± 0.26*	67.75 ± 18.01
7	63.75 ± 3.05	57.57 ± 0.54	58.32 ± 2.49	52.51 ± 24.06*
8	66.75 ± 5.03	58.27 ± 0.75	60.34 ± 0.44*	61.64 ± 28.83*
9	97.27 ± 22.74	74.39 ± 0.46	74.25 ± 0.68	63.60 ± 14.96*
10	73.76 ± 2.85	61.58 ± 0.50	61.40 ± 0.35	50.80 ± 13.72*

* $p < 0.05$.

PPG Signal Reconstruction Performance in Finger Experiment

The performance of the signal reconstruction of the proposed IMAR approach is compared to TD-ICA for the PPG data with an index finger moving left-to-right patterns. The pulse oximeter on the middle finger of the right hand, which was stationary, was used as the reference signal. Since the subjects were directed to produce the motions for 30 s within each 1-min segment, corresponding to 50% corruption by duration, the window length of both clean and corrupted segments were both set as half length of the signal.

TABLE 9. Comparison and statistical analysis of HR estimations from IMAR-reconstructed PPG for 9 different subjects (Walking & Stair Climbing Experiment).

Subject	Head HR (mean ± SD)	Finger HR (Reference) (mean ± SD)	IMAR reconstructed HR (mean ± SD)	ICA reconstructed HR (mean ± SD)
1	62.16 ± 18.96	70.73 ± 5.80	70.55 ± 0.56	77.39 ± 11.90*
2	94.30 ± 20.37	94.40 ± 1.69	95.54 ± 0.86	92.94 ± 9.99*
3	105.53 ± 17.23	120.64 ± 2.98	122.00 ± 1.05	95.67 ± 13.10*
4	95.48 ± 8.37	101.61 ± 3.06	99.89 ± 0.44*	90.89 ± 8.28*
5	82.20 ± 13.07	86.99 ± 3.71	87.71 ± 1.07	82.84 ± 17.96*
6	77.40 ± 12.69	82.48 ± 1.68	81.93 ± 0.48	86.81 ± 12.54*
7	121.02 ± 19.26	107.58 ± 1.51	109.15 ± 0.07	138.62 ± 6.18*
8	86.57 ± 9.85	91.95 ± 6.07	91.73 ± 0.57	80.44 ± 4.61*
9	87.09 ± 6.56	82.55 ± 5.24	84.22 ± 1.93*	104.30 ± 21.43*

* $p < 0.05$.

TABLE 10. Comparison and statistical analysis of SpO_2 estimations from IMAR-reconstructed PPG for 9 different subjects (Walking & Stair Climbing Experiment).

Subject	Head SpO_2 (mean ± SD)	Finger SpO_2 (Reference) (mean ± SD)	IMAR reconstructed SpO_2 (mean ± SD)	ICA reconstructed SpO_2
1	95.70 ± 7.62	99.00 ± 0	97.64 ± 2.50	84.21 ± 1.34*
2	94.55 ± 5.51	95.37 ± 0	96.37 ± 0.99	95.53 ± 1.59
3	91.00 ± 15.58	96.75 ± 0	94.51 ± 0.42*	84.64 ± 4.63*
4	89.61 ± 3.36	99.62 ± 0	102.25 ± 0.65*	87.33 ± 2.67*
5	94.27 ± 8.12	98.00 ± 0.50	97.34 ± 1.45	76.50 ± 1.53*
6	88.50 ± 13.95	96.00 ± 0.31	94.97 ± 4.07*	82.94 ± 1.05*
7	94.92 ± 16.77	98.00 ± 0	100.37 ± 3.15	90.69 ± 8.11*
8	96.11 ± 6.60	97.00 ± 0	98.70 ± 4.16*	96.11 ± 0.39
9	93.78 ± 6.60	98.62 ± 0	95.99 ± 2.39*	89.11 ± 5.03*

Table 8 compares the HR reconstruction results between the IMAR and TD-ICA methods for all 10 subjects. As shown in Table 8, the IMAR reconstructed HR values are not significantly different from the reference HR in 7 out of 10 subjects. However, the TD-ICA's reconstructed HR is significantly different from the reference HR in 8 out of 10 subjects indicating poor reconstruction fidelity.

PPG Signal Reconstruction Performance for the Walking and Stair Climbing Experimental Data

The signal reconstruction of the MNA identified data segments of the walking and stair climbing experiments using our proposed IMAR and its comparison to TD-ICA are provided in this section. Detection of the MNA data segments was performed using the algorithm described in Part I of the companion paper. The reconstructed HR and SpO_2 values using our proposed algorithm and TD-ICA are provided in Tables 9 and 10, respectively. For both HR and SpO_2 reconstruction, the measurements were carried out using PPG data recorded from the head pulse oximeter. The right hand index finger's PPG data was used as HR and SpO_2 references. As shown in Table 9, 7 out of 9 subjects' reconstructed HR values were found to be not significantly different from the reference HR values using our algorithm. While 2 subjects' reconstructed HR values were found to be significantly different than the reference, the differences in the actual HR values are minimal. For TD-ICA's reconstructed HR values, all values deviate significantly from the reference values.

For the reconstructed SpO_2 values, our algorithm again significantly outperforms TD-ICA. All but one subject are not significantly different than the SpO_2 reference values for TD-ICA. For our IMAR algorithm, only 4 out of 9 subjects do not show significant difference from the reference values. Note the zero standard deviation reference SpO_2 values from Massimo's pulse oximeter in 7 out of 9 subjects. This is because Massimo uses a proprietary averaging scheme based on several past values. Hence, it is possible that the significant difference seen with our algorithm in some of the subjects would turn out to be not significant if the averaging scheme were not used. While some of the SpO_2 values from our algorithm are significantly different from the reference, the actual deviations are minimal and they are far less than with TD-ICA.

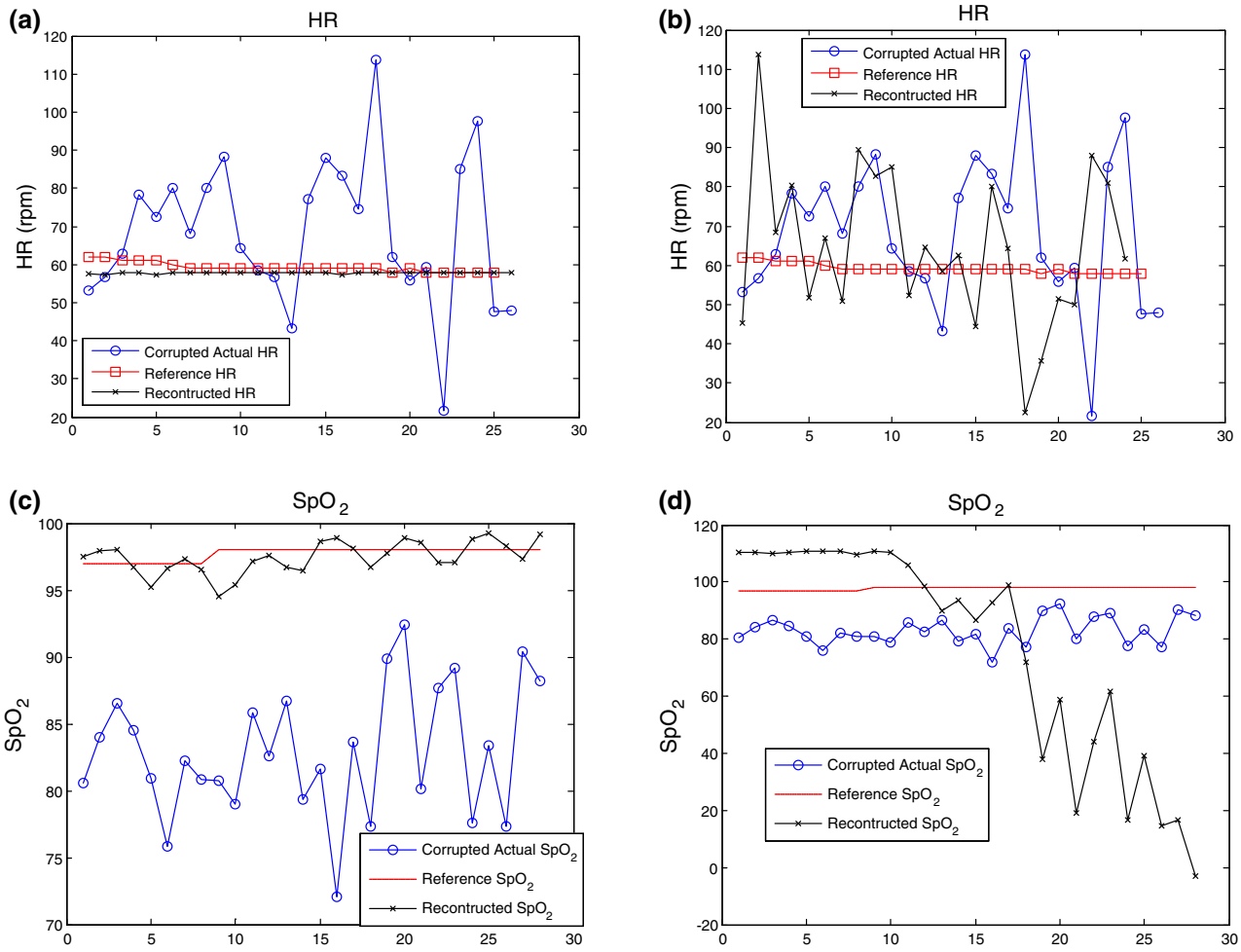


FIGURE 4. (a) HR estimated from IMAR-reconstructed PPG compared to reference and corrupted PPG; (b) HR estimated from ICA-reconstructed PPG compared to reference and corrupted PPG; (c) SpO_2 estimated from IMAR-reconstructed PPG compared to reference and corrupted PPG; (d) SpO_2 estimated from ICA-reconstructed PPG compared to reference and corrupted PPG.

DISCUSSION

In this study, a novel IMAR method is introduced to reconstruct MNA contaminated segments of PPG data. Detection of MNA using a support vector machine algorithm was introduced in the companion paper. The aim of the current paper is to reconstruct the MNA corrupted segments as closely as possible to the non-corrupted data so that accurate HRs and SpO_2 values can be derived. The question is how to reconstruct the MNA data segments when there is no reference signal. To address this question, we use the most adjacent prior clean data segment and use its dynamics to derive the MNA contaminated segment's HRs and oxygen saturation values. Hence, the key assumption with our IMAR technique is that signal's dynamics do not change abruptly between the MNA contaminated segment and its most adjacent prior clean portion of data. Clearly, if this assumption is violated, the IMAR's ability to reconstruct the dynamics of the signal will be compromised. We are currently working on a time-varying IMAR algorithm to address this issue.

There are hosts of algorithms available for MNA elimination and signal reconstruction. Various adaptive filter approaches to remove MNA have been proposed with good results but the test data to fully evaluate the algorithms are either limited or confined to laboratory controlled MNA involving simple finger or arm movements.[11,19,24,25] Moreover, these adaptive filter methods work best when a reference signal is available.

For those methods that do not require a reference signal to remove MNA, there have been many algorithms developed based on variants of the ICA.[16,18,21,25] Most of the ICA-based methods produced reasonably good signal reconstructions of the MNA contaminated data. However, most of these methods were validated on data that were collected using laboratory controlled MNA involving pre-de-

fined simple side-to-side or up-and-down finger and arm movements.[16,18,21,25]

Given that ICA-based methods produced good signal reconstructions of the MNA contaminated data, we have compared our proposed approach to the TD-ICA method as described by Krishnan *et al.*[18] using simulated data, laboratory controlled data as well as daily activity data involving both walking and stair climbing movements. Krishnan *et al.* proposed frequency-domain ICA and compared its performance to the TD-ICA but the improvement is marginal at the expense of higher computational complexity, hence, we used the latter method. Comparison of the performance of our method to TD-ICA was based on reconstruction of HR and SpO_2 values since these measures are currently used by clinicians.

Comparing HR and SpO_2 estimations of the reconstructed signal to the reference measurements using both simulation and experimental data have shown that the proposed IMAR method is a promising tool as the reconstructed values were found to be accurate. The simulation results from noise sensitivity analysis showed that SNR level down to −20 and −15 dB from additive white and colored noise, respectively, can be tolerated well by the application of the proposed IMAR procedure, compared to the SNR values of −10 and −15 dB for the TD-ICA method. Application of the proposed IMAR approach and the TD-ICA to three different sets of experimental data have also shown significantly better signal reconstruction performance with our IMAR algorithm. It is our opinion that ICA is not a good approach for signal reconstruction because its performance suffers from arbitrary scaling and random gain changes in the output signal; these would detrimentally affect the accuracy of SpO_2 estimates.

The use of singular spectrum analysis (SSA) to a single channel EEG recordings to extract high amplitude and low frequency MNA has been previously performed.[26] The main aim of the work by Teixeira *et al.* was to remove the artifacts in EEG signals, hence, an iterative approach to reconstruct the main dynamics of the signal was not implemented. The novelty of our approach is based on the use of SSA combined with an iterative approach to reconstruct the portion of the MNA contaminated data with the most likely true dynamics (i.e., non-MNA contaminated data) of the pulse oximeter signal. In summary, the advantages of the IMAR algorithm is that it can obtain accurate frequency dynamics and amplitude estimates of the signal, hence, the reconstructed HR and SpO_2 estimates should not deviate much from the true uncorrupted values. The disadvantage of the IMAR algorithm is that because it is an iterative approach, the computational complexity is high, hence, in the current form, it is most suitable for offline data analysis. We are not aware of any previous applications of SSA-based algorithms for MNA reconstruction of pulse oximeter data. In conclusion, a scenario where a reference signal is not available to remove the MNA, the proposed IMAR algorithm is a promising new approach to accurately reconstruct HR and SpO_2 values from MNA contaminated data segments.

ACKNOWLEDGMENTS

This work was supported in part by the US Army Medical Research and Materiel Command (USAMRMC) under Grant No. W81XWH-12-1-0541.

REFERENCES

[1]Bishop, G and G. Welch. An Introduction to the Kalman Filter. Tech. Rep, 2006.
[2]Choi, S., A. Cichocki, H. Min Park, and S. Young Lee. Blind source separation and independent component analysis: a review. *Neural Inf. Process. Lett. Rev.* 6:1–57, 2004.
[3]Chon, K. H., S. Dash, and K. Ju. Estimation of respiratory rate from photoplethysmogram data using time-frequency spectral estimation. *IEEE 'Trans. Bio-med. Eng.* 56:2054–2063, 2009.
[4]Comon, Pierre. Independent component analysis, a new concept? *Signal Process.* 36:287–314, 1994.
[5]Diniz, P. Adaptive Filtering: Algorithms and Practical Implementation. Berlin: Springer Science, Business Media L.L.C., 2008.
[6]Elsner, J. B., and A. A. Tsonis. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Berlin: Springer, 1996.
[7]Feleppa, J., and E. J. Mamou. Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds. *Acoust. Soc. Am.* 21:1790–1801, 2007.
[8]Gao, J., C. Zheng, P. Wang. Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis. *Clin. EEG Neurosci.* 41(1):53–59, 2010.
[9]Ghaderi, F., H. R. Mohseni, and S. Sanei. Localizing heart sounds in respiratory signals using singular spectrum analysis. *IEEE Trans. Biomed. Eng.* 58:3360–3367, 2011.
[10]Golyandina, N., V. Nekrutkin, and A. A. Zhigljavsky. Analysis of Time Series Structure: SSA and Related Techniques. London: Taylor & Francis, 2001.
[11]Hamilton, P. S., M. G. Curley, R. M. Aimi, *et al.* Comparison of methods for adaptive removal of motion artifact. In: Computers in Cardiology 2000, pp. 383–386, 2000.
[12]Hassani, H. Singular spectrum analysis: methodology and comparison. *J. Data Sci.* 5:239–257, 2007.
[13]Jolliffe, I. T. Principal Component Analysis. Berlin: Springer, 1986.
[14]Jubran A. Pulse oximetry. In: Applied Physiology in Intensive Care Medicine, edited by G. Hedenstierna, J. Mancebo, L. Brochard, *et al.* Berlin: Springer, pp. 45–48.
[15]Kalman, R. E. A new approach to linear filtering and prediction problems transactions of the ASME. *J. Basic Eng. Ser. D* 1:2, 1960.

[16]Kim, B. S., and S. K. Yoo. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng*. 53:566–568, 2006.
[17]Kloeden, P. E., and E. Platen. Numerical Solution of Stochastic Differential Equations. Berlin: Springer, 1992.
[18]Krishnan, R., B. Natarajan, and S. Warren. Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data. *IEEE Trans. Biomed. Eng*. 57:1867–1876, 2010.
[19]Morbidi, F., A. Garulli, D. Prattichizzo, *et al*. Application of Kalman filter to remove TMS-induced artifacts from EEG recordings. *IEEE Trans. Control Syst. Technol*. 16:1360–1366, 2008.
[20]Ram, M. R., K. V. Madhav, E. H. Krishna *et al*. Use of multi-scale principal component analysis for motion artifact reduction of PPG signals. Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE. pp. 425–430, 2011.
[21]Ram, M. R., K. V. Madhav, E. H. Krishna, *et al*. ICA-based improved DTCWT technique for MA reduction in PPG signals with restored respiratory information. *IEEE Trans. Instrum. Meas*. 62:2639–2651, 2013.
[22]Reddy, K. A., B. George, N. Madhu Mohan *et al*. A novel method of measurement of oxygen saturation in arterial blood. Instrumentation and Measurement Technology Conference Proceedings, 2008 IMTC 2008 IEEE, pp. 1627–1630, 2008.
[23]Reddy, K. A., B. George, N. M. Mohan, *et al*. A novel calibration-free method of measurement of oxygen saturation in arterial blood. *IEEE Trans. Instrum. Meas*. 58:1699–1705, 2009.
[24]Seyedtabaii, S. and L. Seyedtabaii. Kalman Filter Based Adaptive Reduction of Motion Artifact from Photoplethysmographic Signal, Vol. 37. World Academy of Science, Engineering and Technology, 2008.
[25]Sweeney, K. T., T. E. Ward, and S. F. Mcloone. Artifact removal in physiological signals—practices and possibilities. *IEEE Trans. Inf. Technol. Biomed*. 16:488–500, 2012.
[26]Teixeira, A. R., A. M. Tomé, E. W. Lang, *et al*. Automatic removal of high-amplitude artefacts from single-channel electroencephalograms. *Comput. Methods Programs Biomed*. 83:125–138, 2006.
[27]Thakor, N. V., and Y.-S. Zhu. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Trans. Biomed. Eng*. 38:785–794, 1991.
[28]Thompson, B. Canonical Correlation Analysis: Uses and Interpretation. Thousand Oaks: SAGE Publications, 1984.

Sensors **2014**, *14*, 13830-13850; doi:10.3390/s140813830

OPEN ACCESS

sensors

ISSN 1424-8220

www.mdpi.com/journal/sensors

Article

Tracheal Sounds Acquisition Using Smartphones

Bersain A. Reyes, Natasa Reljin and Ki H. Chon *

Department of Biomedical Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA; E-Mails: bareyes@wpi.edu (B.A.R.); nreljin@wpi.edu (N.R.)

* Author to whom correspondence should be addressed; E-Mail: kichon@wpi.edu; Tel.: +1-508-831-4114; Fax: +1-508-831-4121.

Received: 12 June 2014; in revised form: 22 July 2014 / Accepted: 25 July 2014 / Published: 30 July 2014

Abstract: Tracheal sounds have received a lot of attention for estimating ventilation parameters in a non-invasive way. The aim of this work was to examine the feasibility of extracting accurate airflow, and automating the detection of breath-phase onset and respiratory rates all directly from tracheal sounds acquired from an acoustic microphone connected to a smartphone. We employed the Samsung Galaxy S4 and iPhone 4s smartphones to acquire tracheal sounds from $N = 9$ healthy volunteers at airflows ranging from 0.5 to 2.5 L/s. We found that the amplitude of the smartphone-acquired sounds was highly correlated with the airflow from a spirometer, and similar to previously-published studies, we found that the increasing tracheal sounds' amplitude as flow increases follows a power law relationship. Acquired tracheal sounds were used for breath-phase onset detection and their onsets differed by only 52 ± 51 ms (mean ± SD) for Galaxy S4, and 51 ± 48 ms for iPhone 4s, when compared to those detected from the reference signal via the spirometer. Moreover, it was found that accurate respiratory rates (RR) can be obtained from tracheal sounds. The correlation index, bias and limits of agreement were $r^2 = 0.9693$, 0.11 (−1.41 to 1.63) breaths-per-minute (bpm) for Galaxy S4, and $r^2 = 0.9672$, 0.097 (–1.38 to 1.57) bpm for iPhone 4s, when compared to RR estimated from spirometry. Both smartphone devices performed similarly, as no statistically-significant differences were found.

Keywords: respiratory sounds; tracheal sounds; smartphone; respiratory rate; breath-phase; entropy; time-frequency representation

---

1. Introduction

Respiratory sounds vary and they include breath sounds, adventitious sounds, and sounds from the respiratory muscles, excluding voiced sounds during breathing, according to the European Respiratory Society (ERS) Task Force Report [1]. Lung sounds are all respiratory sounds heard or detected over the chest wall or within the chest, including breathing and adventitious sounds detected at this location [1]. Tracheal sounds are those heard or detected over the extrathoracic part of the trachea [1]. In this study, we will concentrate only on the study of tracheal sounds recorded from healthy subjects.

Tracheal sounds exhibit characteristics of noise dynamics with a broad-band spectrum and contain several resonance peaks [2]. Tracheal sounds exhibit well defined inspiratory and expiratory phases and their frequency contents are higher compared to lung sounds [3]. It has been found that inspiratory and expiratory phases have similar frequency contents for tracheal sounds [3,4]. Turbulent flow in upper airways is primarily responsible for generation of tracheal sounds, thus, their characteristics are influenced by airway dimensions [5]. It has been shown that tracheal sounds consist of a dominating local turbulent eddy and a propagating acoustic component with resonances [2]. The relation between airflow, F, and tracheal sound's amplitude, A, has been recognized. A power law of the form $A = kF^{\alpha}$ is considered the typical best fit, where k and α are constants with varying values having been found from different research groups [2,6–9].

The stethoscope remains the most widely used instrument in clinical medicine and its use during auscultation still guides in diagnosis when other pulmonary function testing is not available [10]. However, auscultation with the mechanical stethoscope has limitations [3,11,12]. Namely, it is a subjective process that depends on the skill of the physician [13]; it is limited by the human auditory system [14]; it depends on the stethoscope model used, and the stethoscope itself is more adequate for cardiac auscultation [3]; and the respiratory sounds are not permanently recorded for further analysis.

Over the last decades, some limitations of the stethoscope have been overcome by using Computerized Respiratory Sound Analysis (CORSA). Computerized analysis of respiratory sounds has led to the renaissance of lung auscultation over the last decades but this renewed interest has also produced several different measurement systems by different laboratories [5,15,16]. Fortunately, standardization of CORSA has been addressed, and guidelines for the minimum requirements of CORSA systems have been provided [17,18]. The European Community financed the CORSA project, which explicitly expressed that [19] "one goal of the current technological developments is to combine processing power, storage, miniaturization of components and analysis programed into a small hand-held computerized stethoscope that will provide the clinician with much more useful information than the current simple mechanical stethoscope." Given the need for reliable devices that can record and analyze respiratory sounds in a continuous, non-invasive, and portable fashion, we propose to develop a respiratory sound system based on a smartphone platform.

It is well known that the use of smartphones has become popular and that they are widely available and used for everyday activities including vital sign measurements. By taking advantage of the smartphone's processing power, peripheral noninvasive and cost-effective sensors, and wireless communications capabilities, recent efforts have been made to create various medical applications for self-monitoring. In particular, our research group has made efforts to employ smartphones for health monitoring in the area of cardiac monitoring [20,21].

The development of an inexpensive, reliable, and portable CORSA system would expand the noninvasive diagnostic capabilities of the auscultation procedure when used by general practitioners and pneumologists in the diagnosis of respiratory diseases during the clinical examination. The use of an inexpensive, reliable, and portable system would also enable more health centers to undertake the quantitative analysis of respiratory sounds for diagnosis of respiratory diseases. We hypothesize that a CORSA system that satisfies these characteristics can be implemented using a smartphone.

There have been attempts to develop a portable system for the analysis of respiratory sounds. In particular, the concept of a portable device based on a microcontroller, memory arrays, and liquid crystal displays has been proposed but without sufficient details about implementation results [22]. The concept of a digital stethoscope using a palmtop computer has been also proposed but neither technical detail about the characteristics of the system that guarantee the reliability of the acquired respiratory sound signals nor examples of the acquired signals were provided [23]. Recently, the concept of a smartphone-based asthma monitoring system has been proposed [24,25]. The sounds were processed via custom-designed hardware and the obtained information was wirelessly transmitted to the smartphone to display the processed data. The processed data were then transmitted to a medical database via the Internet [24]. Sounds obtained from Internet sources were transmitted to smartphone and reconstruction techniques were tested [25]. However, like in the previous attempts, no information was provided about the reliability of the system when acquiring real respiratory sounds.

The smartphone-based CORSA system we propose differs from the existing systems in two main ways. First, the signal processing of the acquired respiratory sounds will be performed directly on the smartphone, without the employment of complicated secondary devices with microcontroller-based architectures that increase the energy consumption and the cost of maintenance/upgrade. The smartphone will be used not only to display the respiratory sound signal, but will also control the acquisition stage and perform the signal processing. Second, no wireless communication will be used to transmit the acquired respiratory sounds to the smartphone in order to avoid losses in the quality of the transmitted information and to reduce the energy consumption in the preprocessing stage. The proposed mobile system will be designed to satisfy the standard requirements for a CORSA system and will take advantage of the already available hardware characteristics of the smartphone for the acquisition, visualization, and processing of the respiratory sounds. The proposed system will have the advantage of being non-invasive, low cost, and a portable device which can be used to monitor anytime and anywhere. It should be noted that the developed system was only used to acquire tracheal sounds while the raw recordings were transferred and processed on a computer.

In this paper, the reliability of our proposed smartphone-based system will be tested on the well-known characteristics of the tracheal sounds: well-defined breath phases, similar frequency content for the inspiratory and expiratory phases, and a flow-dependent amplitude relationship. In addition, we aim to detect the breath-phase onsets from the smartphone-acquired tracheal sounds and compare the results with those obtained using the flow signal from a spirometer which is considered the reference. Finally, we will estimate respiration rates from the smartphone-acquired tracheal sound signals and validate them using the respiration rates estimated from the volume changes derived from a spirometer.

2. Material and Methods

2.1. Subjects

Nine healthy non-smoker volunteers (seven males and two females) ages ranging from 23 to 35 years (mean ± standard deviation: 27.9 ± 5.1), weight 68.7 ± 8.1 kg and height 170.7 ± 6.7 cm, were recruited for this study. The study group consisted of students and staff members from Worcester Polytechnic Institute, MA, USA. All volunteers were invited to participate in the study and each consented to be a subject and signed the study protocol approved by the Institutional Review Board of WPI.

2.2. Tracheal Sounds Data Acquisition

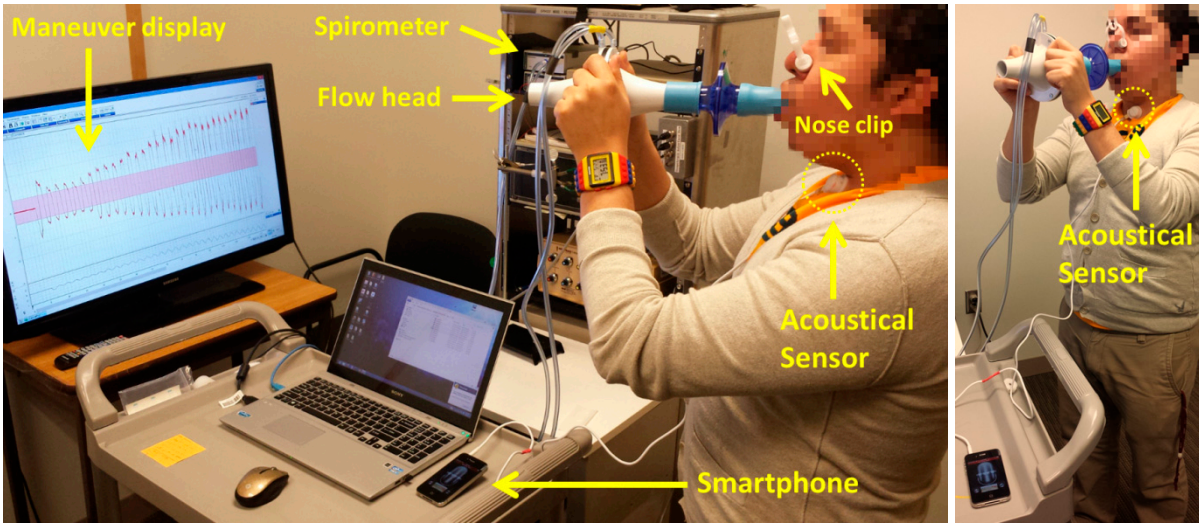
Equipment. Tracheal sounds were acquired using an acoustical sensor composed of a subminiature electret microphone BT-21759-000 (Knowles Electronics, Itasca, IL, USA) encased in a plastic bell. This microphone operates with a voltage supply ranging from 1.3 to 10 V with a low amplifier current drain of 50 μA, provides a flat frequency response between 50 and 3000 Hz, and offers advantages in terms of high durability compared to contact sensors. A light plastic bell was used for air-coupling between the sensor and the recording area on the surface over the trachea. The plastic bell consisted of a conical coupler chamber. This shape provides an efficient transducer of air pressure fluctuations from the skin over the trachea to the microphone [3]. This acoustic sensor was developed by our colleagues at the Metropolitan Autonomous University at Mexico City, Mexico, and had been successfully used for respiratory sound acquisition applications [26,27]. Acoustical sensors of similar characteristics have been found to be adequate for respiratory sound research [17,28,29]. To minimize power line electrical interference, shielded twisted pair cables were used to connect the acoustical sensor to the standard 3.5 mm audio jack in the smartphone. In order to provide impedance matching and to obtain a balance between saturation and sensitivity, a simple voltage divider composed of two resistors of 2.2 kΩ was used before transmitting the recorded tracheal sounds to the smartphone. We cabled to the standard 3.5 mm audio connector to avoid high power consumption or loss of quality due to wireless communication.

Two smartphones were selected for this research: (1) the Galaxy S4 manufactured by Samsung (Samsung Electronics Co., Seoul, South Korea) and running an Android v4.4.2 operating system, and (2) the iPhone 4s manufactured by Apple (Apple, Inc., Cupertino, CA, USA) and running an iOS 6.1 operating system. Selection of the devices was made based on the high market share of each phone's product family, and the dominant combined market share of their operating systems. In addition, each device contains a high-fidelity audio system that satisfies the minimum requirements recommended by the ERS Task Force Report [18]. The tracheal sounds were recorded using the corresponding built-in audio recorder application of each smartphone (Voice Recorder in the Galaxy S4, and Voice Memos in the iPhone 4s) using the predetermined 16-bit per sample and 44.1 kHz sampling rate and saved in the native .m4a format in each device. Recorded audio files were transferred to a personal computer and converted to .wav format preserving the same bits per sample and sampling rate using a conversion software (Free Audio Converter v.5.0.33, DVDVideoSoft Ltd., United Kingdom) and stored for further processing in Matlab (R2012a, The Mathworks, Inc., Natick, MA, USA).

Simultaneously with the tracheal sounds, the airflow was recorded using a spirometer system consisting of a respiratory flow head (MLT1000L, ADInstruments, Inc., Dunedin, New Zealand) connected to a differential pressure transducer to measure airflow (FE141 Spirometer, ADInstruments, Inc., Dunedin, New Zealand). The airflow signal was digitized using a 16-bit A/D converter PowerLab/4SP, ADInstruments, Inc., Dunedin, New Zealand) at 10 kHz sampling rate by using the manufacturer's software (LabChart 7, ADInstruments, Inc., Dunedin, New Zealand). The volume signal was computed online as the integral of the airflow. Prior to each day of recordings, the spirometer system was calibrated using a 3 L calibration syringe (Hans Rudolph, Inc., Shawnee, KS, USA). A new set of disposable filter, reusable mouthpiece, and disposable nose clip (MLA304, MLA1026, MLA1008, ADInstruments, Inc., Dunedin, New Zealand) was given to each subject.

Acquisition protocol. Experiments were performed not in an anechoic chamber but in an office room held quiet. The acoustical sensor was fixed to the neck of the volunteers at the anterior cervical triangle using a double-sided adhesive ring (BIOPAC Systems, Goleta, CA, USA) to avoid pressure variations if hand-placed. Subjects were asked to breathe through a spirometer for approximately 2 min at airflow levels above 0.5 L/s and at maximum of around 2.5 to 3.0 L/s; varying among subjects. The subjects were instructed to breathe first by increasing volumetric flow rates with each breath, and then with decreasing volumetric flow rates with each breath. The airflow was displayed on a 40" monitor placed in front of the subject in order to provide visual feedback. Visual markers were placed between −0.5 to 0.5 L/s and the subjects were instructed to keep the airflow peaks of each respiratory phase outside this boundary area. Initial inspiratory and final expiratory apnea phases of approximately 5 s were acquired in order to record the ambient noise levels. Nose clips were used to clamp the nostrils during the respiratory maneuver. An example of the acquisition protocol is shown in Figure 1. A respiratory maneuver was acquired using each smartphone in a sequential way, where the order of the devices was randomized between subjects.

Figure 1. Tracheal sound recording using a smartphone and simultaneous recording of the airflow signal via spirometer during the respiratory maneuver. The acoustical sensor transmitted the tracheal sound to the smartphone via the standard 3.5 mm audio connector.

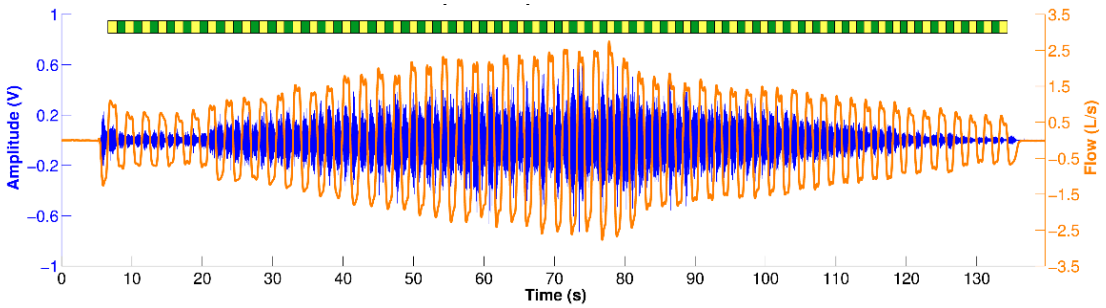


2.3. Data Pre-Processing

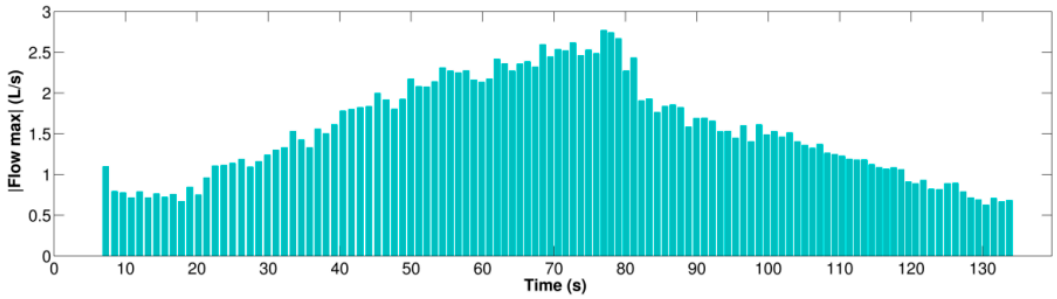
The acquired tracheal sounds were initially down-sampled from 44.1 kHz to 6300 Hz as this frequency still satisfies the Nyquist criteria and reduces the computational burden. Then, the tracheal sounds were digitally filtered using a 4th-order Butterworth filter with a passband between 100 to 3000 Hz to minimize the heart sounds and muscle interferences. The filter was applied in forward and backward scheme to produce zero-phase distortion and minimize the start and end transients. The airflow and the volume signal were down-sampled to 5 kHz and then interpolated to achieve the same sampling frequency of 6300 Hz, and finally they were lowpass filtered at 20 Hz to minimize high frequency components due the interpolation processes that are not related to the respiratory maneuver.

The volume signal was used for automatic segmentation of the inspiratory and expiratory phases by finding its corresponding local maxima and minima during the respiratory maneuver (breath-phase onsets) and by computing the volume slope between both consecutive onsets (positive for inspiration and negative for expiration). Although both the tracheal sounds and the volumetric flow rate were simultaneously recorded, due to different time delays and press start button times of the smartphone, these signals were manually aligned and their durations corrected to the minimum length of both. An example of the filtered, aligned and segmented tracheal sounds and airflow measured via spirometer is shown in Figure 2 for the respiratory maneuver performed by one subject.

Figure 2. Tracheal sounds acquired using a smartphone during the respiratory maneuver. (**a**) Preprocessed tracheal sound (in Volts) aligned with the corresponding spirometer's airflow signal. Yellow and green bars on top indicate the inspiratory and expiratory phases, respectively; (**b**) Corresponding absolute airflow peaks for each respiratory phase of the maneuver.



(**a**)



(**b**)

2.4. Tracheal Sound Amplitude and Airflow Relationship

Although visual inspection of the acquired tracheal sounds indicates that their amplitude increases as airflow increases, and decreases as airflow decreases, we used the information from automatically extracted inspiratory and expiratory phases to quantify this relationship. At each respiratory phase, the peak airflow was found and a time window of 400 ms was created starting at the time instant when the airflow signal reached the upper 10% of the airflow, where it reached its plateau. The tracheal sound segments within these windows were extracted and their corresponding power spectral density (PSD) was computed using the fast Fourier transform (FFT) with $NFFT = 1024$. The PSD of the initial apnea period was also computed and subtracted from each PSD of the tracheal sounds segment. The area under the curve of the resulting PSD was computed and regarded as the amplitude of the tracheal sound for that corresponding respiratory breath-phase. For each subject, the inspiratory/expiratory amplitudes were normalized by dividing them by the average inspiration/expiration amplitude [2]. Finally, for each subject, the best fitting curve of the form $A = kF^{\alpha}$ was computed separately for the inspiratory and expiratory phases.

2.5. Breath-Phase Onset Detection Using Tracheal Sounds

Tracheal sounds acquired with the smartphones were used to estimate the breath-phase onset via the Shannon entropy approach. The Shannon entropy of tracheal sounds has been used as a method for estimating the airflow [30,31]. The Shannon entropy (SE) of a random signal with probability density function (pdf) p is defined as

$$SE(p) = -\sum_{i=1}^{N} p_i \cdot log(p_i) \quad (1)$$

where N is the number of outcomes of the random variable with pdf p. The SE is used to quantify the uncertainty or irregularity of the process [32]. It has been found that the entropy values quantify the standard deviation and correlation properties of the signal where the individual weight contributions are not trivial to separate [33,34].

As proposed for the airflow estimation from tracheal sounds, we applied the Shannon entropy in a moving window scheme as follows. First, the recorded tracheal sounds were sequestered into 25 ms windows with 50% overlap between successive windows. For each of the resulting windows, the Shannon entropy was computed. One way to estimate the pdf p is to use the histogram. However, due to the low number of samples within each overlapping window ($n = 157$ samples) its accuracy would be low. Instead, the pdf of each windowed tracheal sound was computed using the Parzen-window density estimation method with a Gaussian kernel [35,36]. This non-parametric method estimates the pdf p of the random sample x from which the sample was derived, by superposing window functions placed at each of n observations and determining how many observations x_i fall within the specified window h, *i.e.*, the contribution of each observation x_i within this window h. Then, the pdf is estimated as the sum of the total of the contributions from the observations to this window, and the Parzen-window estimate $\hat{p}$ is given by

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - x_i}{h}\right) \quad (2)$$

where $h > 0$ is the window width of the kernel K, which is typically a pdf itself. When a Gaussian kernel is used, the Parzen-window estimate becomes

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\sqrt{2\pi}}exp\left(-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right) \quad (3)$$

where h is the standard deviation of the Gaussian kernel, and was set to [37]

$$h = 1.06 \cdot SD(x) \cdot n^{-\frac{1}{5}} \quad (4)$$

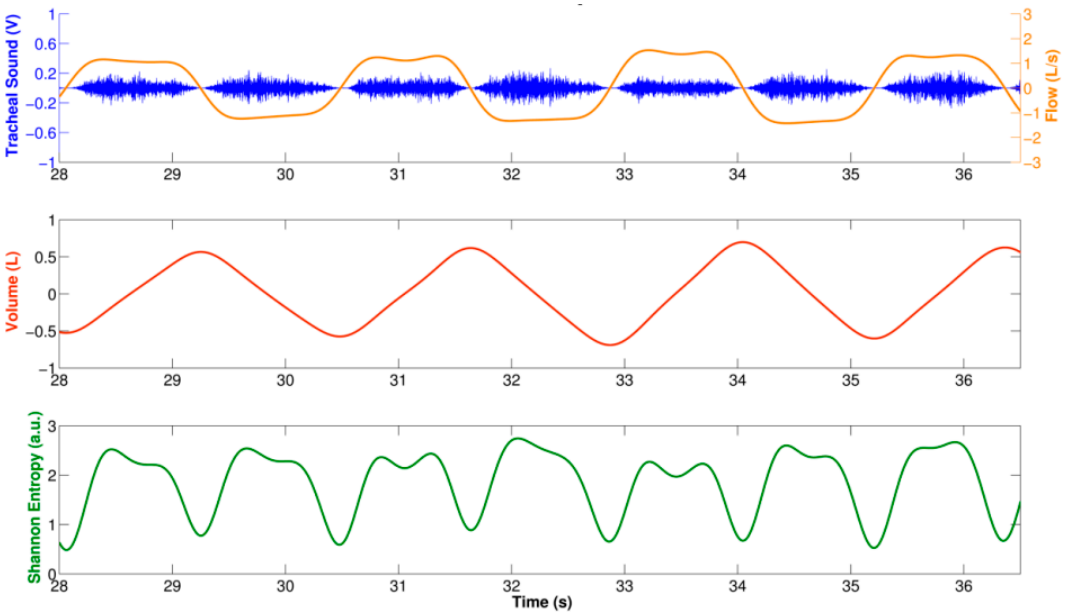
with $SD(\cdot)$ being the standard deviation of the windowed tracheal sound.

The SE estimated from each windowed tracheal sound was assigned to the middle time point of the window, and was interpolated using cubic spline in order to recover the original duration of the tracheal and volumetric flow rate signals. Figure 3 shows an example of the computed SE using the described approach for a tracheal sound segment acquired using a smartphone. Observe that this SE signal from a smartphone resembles a rectified airflow signal, as has previously been found when the SE of tracheal sounds are used for airflow estimation purposes [30,31]. In order to estimate the breath-phase onset, the SE signal was inverted and the corresponding local maxima were automatically detected. First, the SE signal was normalized between [0–1] and down-sampled to 7.875 Hz. The PSD of the down-sampled SE signal was computed with Welch's modified periodogram method with a Hamming window, with 50% overlap, and *NFFT* = 512 bins. The peak of the PSD and its corresponding frequency f_{peak} were found. The local maxima of the SE signal were found and all those maxima that did not satisfy the threshold values criteria were removed. The amplitude threshold was set to $thr_1 = 0.1$, and the time threshold was set to $thr_2 = 0.5 * 1/f_{peak}$. Finally, the corresponding time onsets computed from the down-sampled SE were mapped to the closest point of the original SE time series which had a time resolution of $\Delta_t \approx 0.159\,ms$ given the sampling frequency $f_s = 6300\,Hz$. The detected breath-phase onsets from tracheal sounds acquired from each smartphone were compared to those computed from volume.

2.6. Instantaneous Respiratory Rate Estimation Using Tracheal Sounds

As previously stated, the SE of the tracheal sounds resembles the rectified airflow signal. This SE has two lobes for each breathing cycle indicated by the volume signal; see Figure 3. We took advantage of this fact to estimate the instantaneous respiratory rate from tracheal sounds acquired with the smartphone. In particular, we employed a joint time-frequency representation (TFR) approach. In general, a TFR allows one to analyze which frequencies of a signal under study are present at a certain time, *i.e.*, a TFR describes the energy density of a signal simultaneously in the time and frequency domains [38]. This characteristic is useful when analyzing signals whose frequency content varies with time, as is the case for the respiratory rate.

Figure 3. Shannon Entropy of the tracheal sounds acquired using a smartphone. **Top**: Segment of tracheal sound and corresponding airflow from spirometer (positive lobes are inspirations and negative lobes are expirations); **Middle**: Volume signal obtained with the spirometer as the integral of the flow; **Bottom**: Shannon entropy of tracheal sound. Observe that local minima of the Shannon entropy are obtained around the onset of each respiratory phase.



The most widely-used TFR in the respiratory sounds field is the spectrogram (SP) given by the magnitude square of the short time Fourier transform (STFT) [4,10]. The idea behind the SP is that in order to study the properties of the signal s around time t, the original signal around that time is emphasized but it is suppressed at other times by multiplying by a window function $w(t)$ centered at t, to produce a modified signal $s_t(\tau)$ given by [38,39]

$$s_t(\tau) = s(\tau)w(\tau - t) \tag{5}$$

where the modified signal is a function of two times, the fixed time t of interest, and the time τ. The window function allows the modified signal to satisfy

$$s_t(\tau) = \begin{cases} s(\tau) & for\ \tau\ close\ to\ t \\ 0 & for\ \tau\ far\ away\ from\ t \end{cases} \tag{6}$$

Given that the modified signal emphasizes the original signal around time t, its Fourier transform reflects the frequency distribution around that time

$$\begin{aligned} S_t(\omega) &= \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s_t(\tau) d\tau \\ S_t(\omega) &= \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau) w(\tau - t) d\tau \end{aligned} \tag{7}$$

and hence the name of STFT. The corresponding spectral density at time t is given by

$$SP(t,\omega) = |S_t(\omega)|^2 = \left| \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau) w(\tau - t) d\tau \right|^2 \quad (8)$$

where a spectrum is obtained at each time instant and the total of that spectrum is the time-frequency distribution of the original signal $SP(t,\omega)$. This distribution has received different names depending on the application field, *e.g.*, respirosonogram in the respiratory sounds field [10], but the most common is simply spectrogram.

The SP was applied to the volume signal and to the SE of the acquired tracheal sounds. Due to the very low frequency content of the respiratory rate compared to the original sampling frequency, both signals were down-sampled to 7.875 Hz. The SP was computed using *NFFT* = 512 frequency bins, and a Hamming window of 10 s duration. The resulting TFR was normalized between [0–1] and at each time instant, the maximum peak was computed around the central frequency of the whole signal and the corresponding frequency vector was extracted. Due to the discussion mentioned above, the frequency vector extracted from the volume was regarded as the reference instantaneous respiratory frequency, while the half of the frequency vector extracted from the SE signal corresponded to the instantaneous frequency estimated from each smartphone. All instantaneous respiratory frequencies were converted from Hz to breaths-per-minute (bpm).

For each smartphone, three performance indices were computed for the instantaneous respiratory rate (IRR) of each subject. The first index corresponds to the cross-correlation coefficient ρ between the IRR obtained with the corresponding smartphone and the one obtained from the volume from spirometer given by

$$\rho = \frac{\sum_{i=1}^{N} IRR_{volume}(i) \cdot IRR_{smartphone}(i)}{\sqrt{\sum_{i=1}^{N} (IRR_{volume}(i))^2 \cdot \sum_{i=1}^{N} \left(IRR_{smartphone}(i)\right)^2}} \quad (9)$$

where IRR_{volume} represents the instantaneous respiratory rate obtained from the volume, $IRR_{smartphone}$ the corresponding IRR estimated from the tracheal sound acquired with the iPhone 4s or Galaxy S4 smartphone, and N is the length of the time vector of the signal. Observe that if the IRR_{volume} and $IRR_{smartphone}$ are the same, the value of ρ is unitary. Therefore, ρ values close to 1 reflect a good estimation performance. The remaining two indices computed were the root-mean-squared error $RMSE$, and the normalized root-mean-squared error $NRMSE$, given by

$$RMSE = \sqrt{\frac{\left(\sum_{i=1}^{N} IRR_{volume}(i) - IRR_{smartphone}(i)\right)^2}{N}} \quad (10)$$

$$NRMSE = \frac{RMSE}{mean(IRR_{volume})} \; x\, 100\% \quad (11)$$

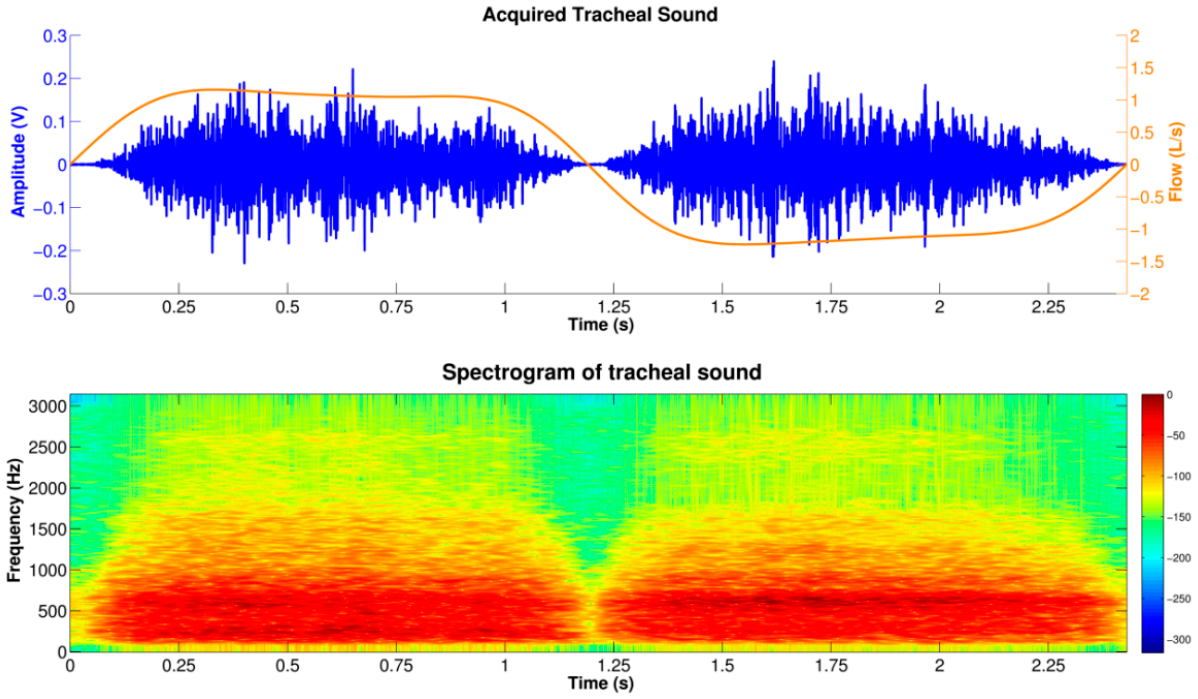
respectively.

3. Results and Discussion

The tracheal sound signals acquired using both the Galaxy S4 and the iPhone 4s showed a temporal intensity variation related to the airflow during the respiratory phases as shown in Figure 4. The TFR of the smartphone-acquired tracheal sounds (bottom panel of Figure 4) shows characteristics of broad band noise where both inspiratory and expiratory phases have their main frequency components not higher than 1.5 kHz, with a sharp drop in power around 800 Hz, which is in agreement with other studies [40]. In addition, both respiratory phases have similar frequency content for similar airflow peaks and a silent period separating both phases could also be observed from both the tracheal sound waveform as well as its TFR. These results are in agreement with the findings reported in the literature when using CORSA systems [3,4,10].

In the next subsections we present the results obtained for both smartphones for the tracheal sound's amplitude with airflow, the breath-phase onset detection, and the respiratory rate estimation.

Figure 4. Time-frequency characteristics of the tracheal sounds acquired using the smartphone during one respiratory cycle. **Top**: Tracheal sound waveform together with its corresponding airflow signal (positive and negative lobes indicate the inspiration and expiration, respectively); **Bottom**: Time-frequency representation of the acquired tracheal sound computed via the spectrogram using a 100 ms Hamming window. Red/blue color in the color map indicates high/low intensity in decibels.



3.1. Tracheal Sound Amplitude and Airflow Relationship

A representative example of the curve fitting of the tracheal sound *versus* airflow acquired with an iPhone 4s and spirometer, respectively, for the inspiration and expiration phases is shown in Figure 5. We observe that during the inspiratory and expiratory phases, the increasing tracheal sounds'

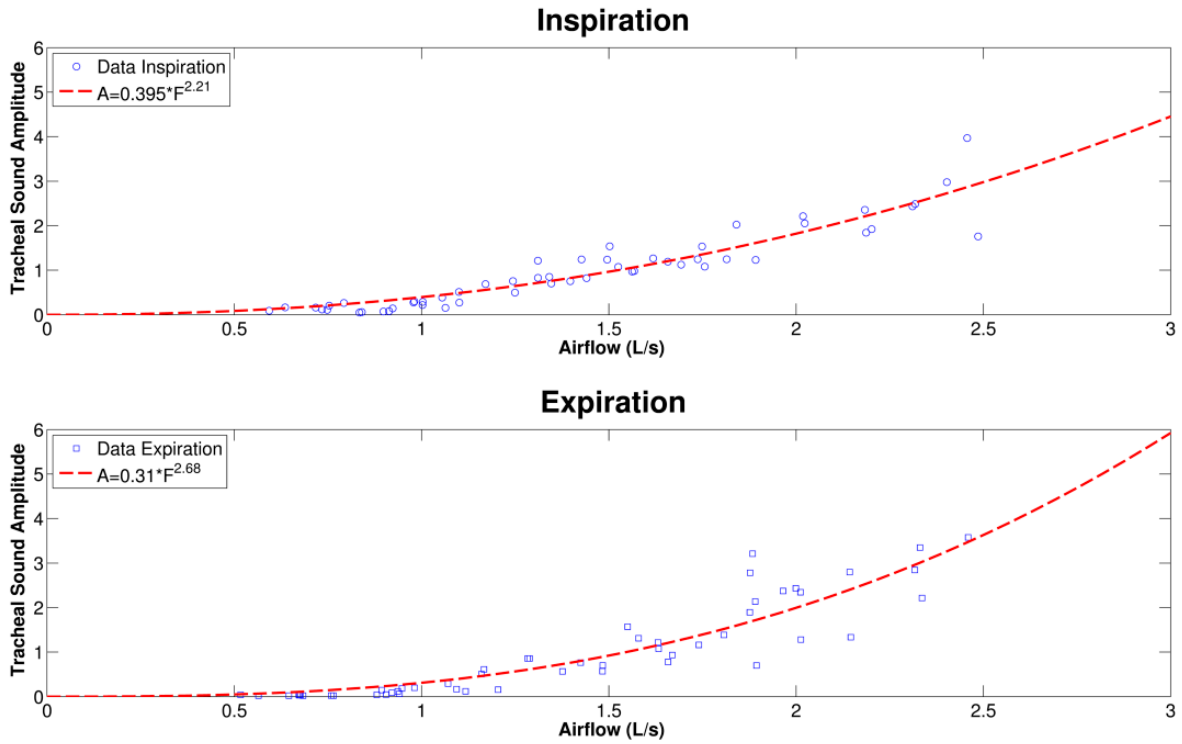
amplitudes as flow increases follow a power law relationship. The results of the power law model fitting parameters for the smartphone-acquired tracheal sounds' amplitude and airflow are presented in Table 1 for each respiratory phase and the two models of smartphones. The mean values of the exponent α were between $\alpha = 1$ and $\alpha = 3$ for both smartphones. It is worth mentioning that different values of the exponent have been found in different studies, ranging from $\alpha = 1$ [7], $\alpha = 2$ [9], $\alpha = 3$ [8], and values in between this range [2,6]. No statistically significant differences were found between the power law parameters obtained from the Galaxy S4 and the iPhone 4s smartphones with a two-tailed paired t-test with $p < 0.05$ considered as statistically significant (SPSS Statistics 17, IBM Corporation, Armonk, NY, USA).

Table 1. Results of the smartphone-acquired tracheal sounds amplitude and airflow relationship using a model of the form $A = kF^{\alpha}$ ($N = 9$ subjects).

Respiratory Phase	Parameter	Galaxy S4	iPhone 4s
Inspiration	k	0.450 ± 0.218	0.371 ± 0.197
	α	2.380 ± 1.077	2.686 ± 0.959
Expiration	k	0.523 ± 0.181	0.349 ± 0.162
	α	1.939 ± 0.900	2.632 ± 0.711

Values presented as mean ± standard deviation.

Figure 5. Example of smartphone-acquired tracheal sounds amplitude as a function of airflow during the inspiratory and expiratory phases for one subject. Red dashed lines correspond to the best fit curves of the form $A = kF^{\alpha}$.



3.2. Breath-Phase Onset Detection Using Tracheal Sounds

Table 2 summarizes the results obtained for the breath-onset detection using each smartphone for all volunteers. The absolute time difference between the reference breath-phase onsets from the volume via the spirometer and the estimated breath-phase onsets from the SE of the acquired tracheal sound, $|\Delta_{onset}|$, was computed. In addition, the total number of true breath-phase onsets computed from the volume is presented together with the corresponding extra and missed breath-phase onsets computed from the tracheal sounds. Note that the total number of true onsets is not the same for both smartphones given that different maneuver trials were performed for each subject. An example of the breathing onset using the iPhone 4s smartphone is shown in Figure 6. As shown, the Δ_{onset} is not consistently positive or negative. The distribution of time onsets was computed via the histogram and is shown in Figure 7 for each smartphone. We found that on average the breath-phase onsets $|\Delta_{onset}|$ detected with the smartphones have a time difference of approximately 50 ms from the onsets detected from the volume. Since some onsets were detected before or after the reference onsets, overall these Δ_{onset} values compensate and the mean onsets differ 9 ms and 14 ms for Samsung S4 and iPhone 4s, respectively. A two-tailed two-sample t-test was performed for Δ_{onset} and $|\Delta_{onset}|$ obtained from both smartphones for the total number of onsets. No statistically significant differences ($p > 0.05$) were found between Δ_{onset} and $|\Delta_{onset}|$ computed from the Galaxy S4 and iPhone 4s.

Table 2. Results of the breath-phase onset detection using smartphone-acquired tracheal sounds in comparison to those detected from volume signal ($N = 9$ subjects).

Parameter		Galaxy S4	iPhone 4s
$\|\Delta_{onset}\|$	$[s]$	0.052 ± 0.051	0.051 ± 0.048
Total onsets		767	854
Extra onsets		12	5
Missed onsets		5	6

Values presented as mean ± standard deviation.

Figure 6. Example of breath-phase onset detection using the tracheal sounds acquired using a smartphone. Solid red lines indicate the breath-phase onsets detected using the volume signal from the spirometer. Dashed blue lines indicate the breath-phase onsets detected using only the information from the acquired tracheal sound.

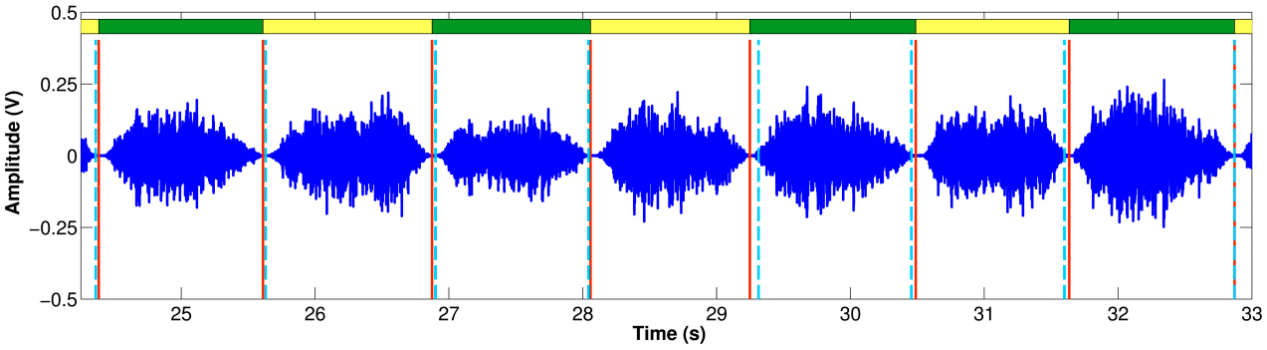
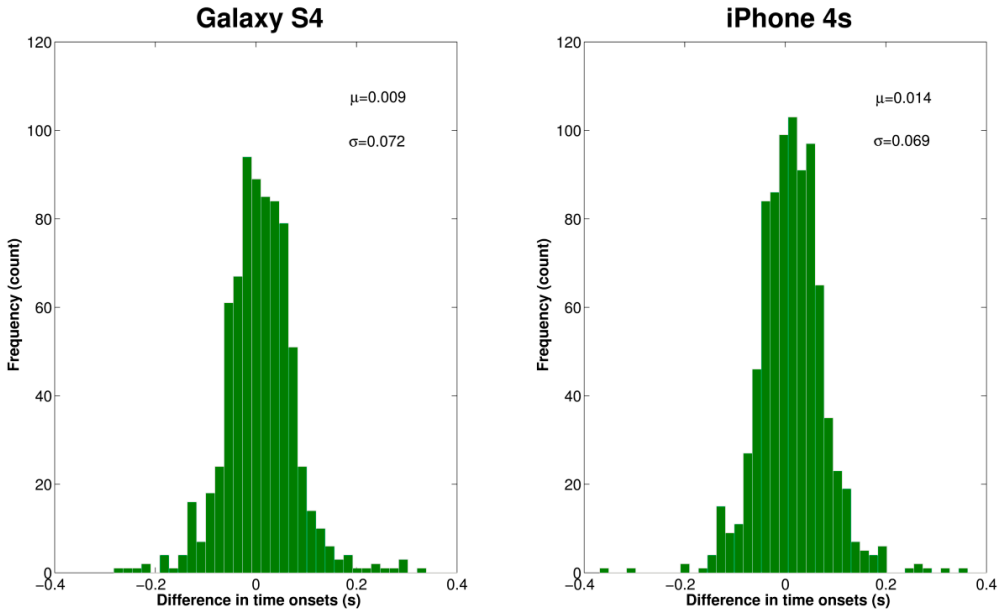


Figure 7. Distribution of the time differences of breath-phase onsets (Δ_{onset}) detected using the volume signal from the spirometer and breath-phase onsets detected using tracheal sounds acquired with the smartphones.



3.3. Instantaneous Respiratory Rate (IRR) Estimation Using Tracheal Sounds

The IRR estimation process using the spectrogram is illustrated in Figure 8 for a tracheal sound acquired using the iPhone 4s. It can be seen that the main frequency of the SE of the tracheal sound (Figure 8b), is located at twice the main frequency of the volume (Figure 8a), which is considered as reference, as the SE resembles a rectified airflow signal. At each time instant, the frequency at which the maximum energy of the TFR occurs was extracted from the corresponding spectrogram (white dashed lines superimposed on TFRs). Comparison of the estimated instantaneous frequencies of the SE of tracheal sound and volume of the spirometer is shown in Figure 8c. In most cases, we found that the discrepancies were more notable at the beginning and the end of the signal where the airflow levels were lower which in turn provided tracheal sound signals with small amplitudes. These are reflected as dispersion points in Figure 9. Table 3 summarizes the IRR results for all the subjects in terms of the performance indices for each smartphone. For both smartphones we found high cross correlation coefficients between the IRR estimated from tracheal sounds and volume. This is also reflected in the regression lines in Figure 9a,c for the Galaxy S4 (r^2 = 0.9693) and iPhone 4s (r^2 = 0.9672), respectively. High linear correlation has been also found between a tracheal acoustical method and pneumotachometer (r^2 = 0.98) [41]. A two-tailed paired t-test was performed for each performance index obtained from both smartphones for all subjects. For all performance indices, no statistically significant differences ($p > 0.05$) were found between the results from Galaxy S4 and iPhone 4s for the estimation of IRR considering the volume from spirometer as reference. The regression lines and the Bland-Altman plots between the estimated instantaneous respiration rate from tracheal sounds and the reference instantaneous respiration rate from volume signals are presented in Figure 9 for the Galaxy S4 and iPhone 4s smartphones. Compared to the spirometer, the bias ± 1.96SD and 95% limits of agreement were 0.11 ± 1.52 bpm and −1.41 to 1.63 bpm for the Galaxy S4, and 0.097 ± 1.47 bpm and −1.38 to 1.57 bpm for the iPhone 4s.

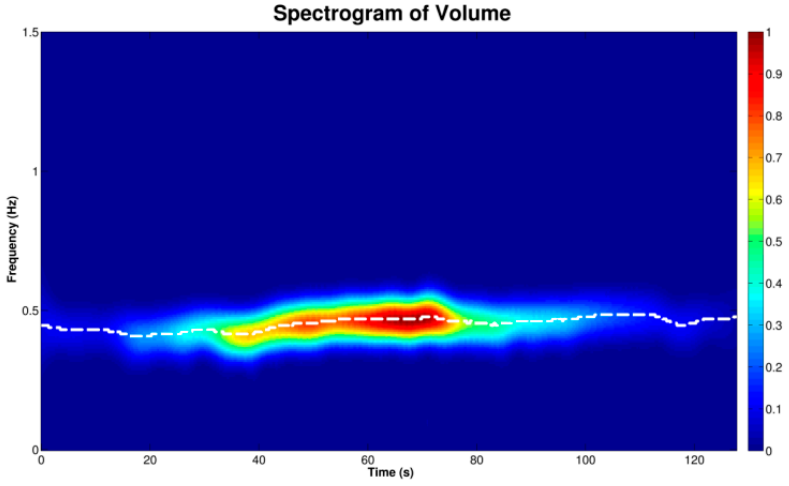
Similar correlations and limits of agreement have been reported for a commercial device in post-anesthesia patients in comparison to capnography [42].

Table 3. Results of the instantaneous respiratory rate estimation using tracheal sounds acquired with the smartphones in comparison to those from volume signals ($N = 9$ subjects).

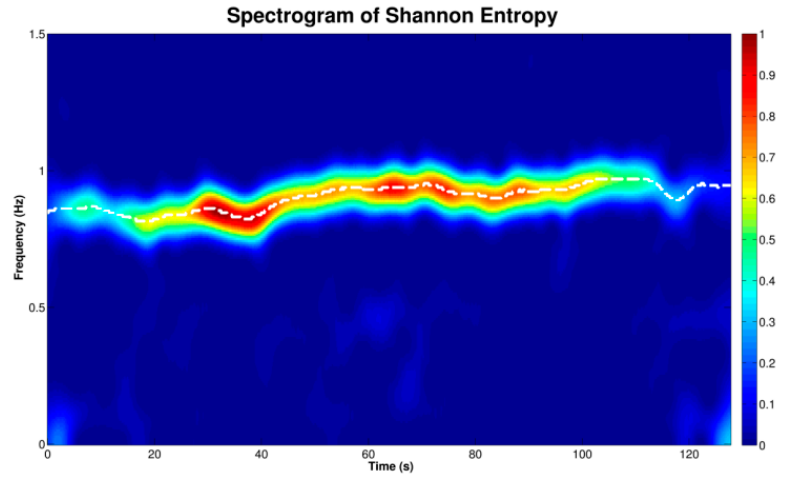
Parameter		Galaxy S4	iPhone 4s
ρ	[*unitless*]	0.9994 ± 0.0004	0.9995 ± 0.0004
RMSE	[*bpm*]	0.731 ± 0.2878	0.700 ± 0.367
NRMSE	[*%*]	3.218 ± 1.297	2.957 ± 1.322

Values presented as mean ± standard deviation.

Figure 8. Estimation of the instantaneous respiratory rate using tracheal sounds acquired with a smartphone. (**a**) Spectrogram of the volume signal from the spirometer; (**b**) Spectrogram of the Shannon entropy of tracheal sound acquired with a smartphone. Observe that the main frequency content of the entropy signal is located at twice the frequency of that from the volume signal. White dashed lines indicate the maximum peak at each time instant; (**c**) Instantaneous respiratory rate computed from corresponding spectrograms of volume and Shannon entropy of tracheal sound.

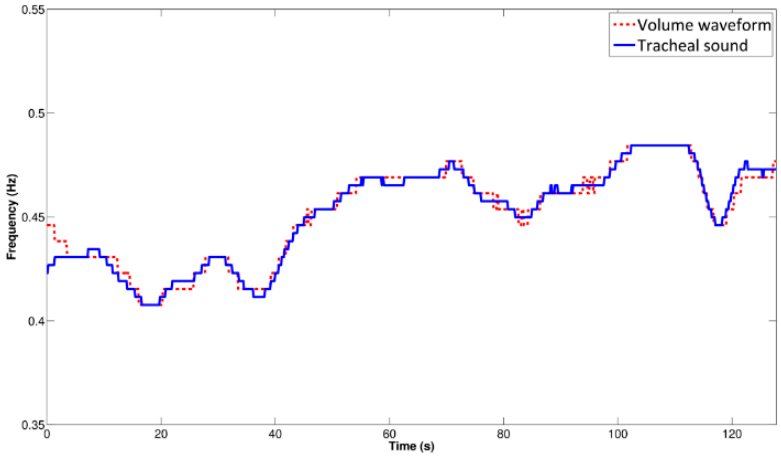


(**a**)



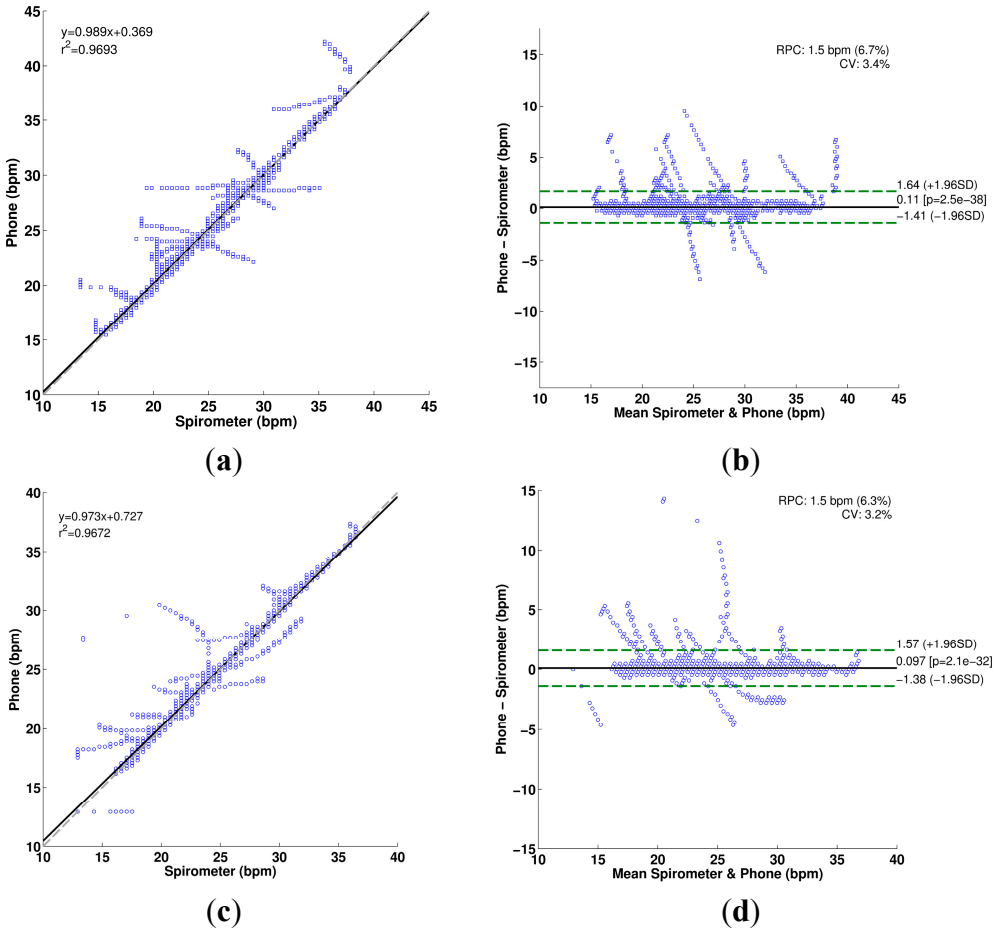
(**b**)

Figure 8. *Cont.*



(**c**)

Figure 9. Comparison of instantaneous respiratory rate estimated from tracheal sounds acquired with smartphones and estimated from volume signals (*N*=9 subjects). (**a**) Regression line for estimation from Galaxy S4; (**b**) Bland-Altman plot for estimation from Galaxy S4; (**c**) Regression line for estimation from iPhone 4s; (**d**) Bland-Altman plot for estimation from iPhone 4s. In regression plots, the grey dashed line indicates the identity line and the solid black the regression line. In Bland-Altman plots, the solid black line indicates the bias while the dashed green lines indicate the 95% limits of agreement.



(**a**) (**b**)

(**c**) (**d**)

4. Conclusions

In this paper, we propose the use of smartphones to develop a CORSA system that satisfies the current standards in the field. In particular, we employed two market-leading smartphones, the Galaxy S4 and iPhone 4s, and specifically-designed respiratory acoustical sensors for the acquisition of tracheal sounds. We obtained tracheal sounds from healthy volunteers in airflow controlled conditions from 0.5 to 2.5 L/s in a quiet room, but not an anechoic chamber.

The relationship between amplitude of tracheal sounds and airflow has been shown to be useful for respiratory health monitoring [30,31,43]. Tracheal sounds have been used for estimation of the airflow and volume in a non-invasive way [30,31,43]. The tracheal sounds have been used to estimate ventilation parameters by first estimating the airflow and then integrating this signal to estimate the volume [43]. Several features have been used to estimate the airflow from the tracheal acoustical information, and the Shannon entropy of the tracheal sounds was found to provide better performance compared to other models based on the signal envelope and average power [30,31]. In this work, we found that smartphone-acquired tracheal sounds' amplitude is proportional to the airflow from a spirometer in a power law relationship which is in agreement to prior studies [2,6–9]. The power law models found for the inspiratory phase were $A = (0.450 \pm 0.218)F^{(2.380 \pm 1.077)}$ for the Galaxy S4, and $A = (0.371 \pm 0.197)F^{(2.686 \pm 0.959)}$ for the iPhone 4s, while for the expiratory phase they were $A = (0.523 \pm 0.181)F^{(1.939 \pm 0.900)}$ for the Galaxy S4, and $A = (0.349 \pm 0.162)F^{(2.632 \pm 0.711)}$ for the iPhone 4s.

Apnea monitoring and automatic breath-phase detection have been other applications of tracheal sounds analysis [44,45]. In particular, information from the logarithm of the variance of the tracheal sounds was used as a way to detect the breath-phase onset which becomes a crucial part in an automatic acoustical system. Towards this goal, we tested the ability of the smartphone-acquired tracheal sounds to detect the breath-phase onsets, as this processing stage is important when the acoustical approach is used for airflow measurement and automatic breath-phase classification. Our results indicate that on average the onsets estimated from the smartphone-acquired tracheal sounds differ by only 52 ± 51 ms for Galaxy S4, and 51 ± 48 ms for iPhone 4s, from the corresponding onsets detected from the spirometer reference signal.

Estimation of the respiratory rate using an acoustical approach has recently gained popularity in clinical settings. As a vital sign, respiration rate can be used to predict serious clinical events [46]. In particular, continuous monitoring of breathing status becomes relevant to identify and predict risk situations both inside and outside clinical settings. Current clinical continuous monitoring methods include qualified human observation, impedance pneumography, and capnography monitoring. However, these methods have disadvantages, *e.g.*, low tolerance of the patient for using the nasal cannula, or leaks around this cannula in capnography. As an alternative, respiratory rate estimation based on tracheal sound has been proposed [41]. Recently, a commercial device that monitors the respiratory rate via tracheal sounds was introduced for clinical settings (Masimo Rainbow SET® Acoustic Monitoring, Masimo Corp., Irvine, CA, USA). The accuracy of this device has been tested against capnography and good correlation has been found between both methods [42,47]. However, there is still a need for a small and discrete device for everyday use, able to estimate the respiration rate in a continuous and non-invasive way outside the clinical setting [48]. Towards addressing this need, we found good correlation between the smartphone-based respiratory rate estimates and the

spirometer-based ones ($r^2 \approx 0.97$), as well as 95% limits of agreement ranging approximately from −1.4 to 1.6 bpm for subjects breathing in a range from 15 to 35 bpm. Overall we did not find statistically significant differences between the results from the Galaxy S4 and iPhone 4s devices.

By employing smartphone devices we were able to reproduce major findings in the tracheal sounds field obtained with conventional CORSA systems. We foresee that efforts similar to the one performed in this study would result in a reliable, low-cost, and easy-to-upgrade portable system that could aid not only general practitioners but also serve as on-demand health monitors outside clinical settings. In addition, systems with such characteristics would aid in the acquisition of large-sample studies in locations not easily accessible nowadays with the currently-used CORSA systems.

Our future work includes implementation of the presented signal processing techniques into applications on the smartphone operating systems, *i.e.*, Android and iOS, which will govern the acquisition, processing and display of the tracheal sounds information.

Acknowledgments

This work is supported in part by the US Army Medical Research and Material Command (US-AMRMC) under grant No. W81XWH-12-1-0541.

The authors would like to thank Professors Sonia Charleston-Villalobos, Tomas Aljama-Corrales, and Ramon Gonzalez-Camarena for introducing a microphone sensor used in this research work.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Sovijarvi, A.R.A.; Dalmasso, F.; Vanderschoot, J.; Malmberg, L.P.; Righini, G.; Stoneman, S.A.T. Definition of terms for applications of respiratory sounds. *Eur. Respir. Rev.* **2000**, *10*, 597–610.
2. Beck, R.; Rosenhouse, G.; Mahagnah, M.; Chow, R.M.; Cugell, D.W.; Gavriely, N. Measurements and Theory of Normal Tracheal Breath Sounds. *Ann. Biomed. Eng.* **2005**, *33*, 1344–1351.
3. Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory Sounds: Advances beyond the Stethoscope. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987.
4. McKusick, V.A.; Jenkins, J.T.; Webb, G.N. The acoustic basis of the chest examination; studies by means of sound spectrography. *Am. Rev. Tuberc.* **1955**, *72*, 12–34.
5. Dalmay, F.; Antonini, M.T.; Marquet, P.; Menier, R. Acoustic properties of the normal chest. *Eur. Respir. J.* **1995**, *8*, 1761–1769.
6. Gavriely, N.; Cugell, D.W. Airflow effects on amplitude and spectral content of normal breath sounds. *J. Appl. Physiol.* **1996**, *80*, 5–13.
7. Leblanc, P.; Macklem, P.T.; Ross, W.R. Breath sounds and distribution of pulmonary ventilation. *Am. Rev. Respir. Dis.* **1970**, *102*, 10–16.
8. Olson, D.; Hammersley, J. Mechanisms of Lung Sound Generation. Semin. *Respir. Crit. Care Med*. **1985**, *6*, 171–179.

9. Shykoff, B.E.; Ploysongsang, Y.; Chang, H.K. Airflow and Normal Lung Sounds. *Am. Rev. Respir. Dis.* **1988**, *137*, 872–876.
10. Pasterkamp, H.; Carson, C.; Daien, D.; Oh, Y. Digital respirosonography. New images of lung sounds. *Chest* **1989**, *96*, 1405–1412.
11. Forgacs, P. The functional basis of pulmonary sounds. *Chest* **1978**, *73*, 399–405.
12. Charbonneau, G.; Ademovic, E.; Cheetham, B.M.G.; Malmberg, L.P.; Vanderschoot, J.; Sovijarvi, A.R.A. Basic techniques for respiratory sound analysis. *Eur. Respir. Rev.* **2000**, *10*, 625–635.
13. Brooks, D.; Thomas, J. Interrater Reliability of Auscultation of Breath Sounds among Physical Therapists. *Phys. Ther.* **1995**, *75*, 1082–1088.
14. Piirila, P.; Sovijarvi, A.R. Crackles: Recording, analysis and clinical significance. *Eur. Respir. J.* **1995**, *8*, 2139–2148.
15. Earis, J.E.; Cheetham, B.M.G. Current methods used for computerized respiratory sound analysis. *Eur. Respir. Rev.* **2000**, *10*, 586–590.
16. Mussell, D.M.J. The need for standards in recording and analysing respiratory sounds. *Med. Biol. Eng. Comput.* **1992**, *30*, 129–139.
17. Cheetham, B.M.G.; Charbonneau, G.; Giordano, A.; Helisto, P.; Vanderschoot, J. Digitization of data for respiratory sound recordings. *Eur. Respir. Rev.* **2000**, *10*, 621–624.
18. Vannuccini, L.; Earis, J.E.; Helisto, P.; Cheetham, B.M.G.; Rossi, M.; Sovijarvi, A.R.A.; Vanderschoot, J. Capturing and preprocessing of respiratory sounds. *Eur. Respir. Rev.* **2000**, *10*, 616–620.
19. Earis, J.E.; Cheetham, B.M.G. Future perspectives for respiratory sound research. *Eur. Respir. Rev.* **2000**, *10*, 641–646.
20. Lee, J.; Reyes, B.A.; McManus, D.D.; Mathias, O.; Chon, K.H. Atrial Fibrillation Detection Using an iPhone 4S. *IEEE Trans. Biomed. Eng.* **2013***, 60*, 203–206.
21. Scully, C.G.; Lee, J.; Meyer, J.; Gorbach, A.M.; Granquist-Fraser, D.; Mendelson, Y.; Chon, K.H. Physiological Parameter Monitoring from Optical Recordings with a Mobile Phone. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 303–306.
22. Liu, G.; Chen, S.; Zhang, J.; Cheng, J.; Wu, S. The development of a portable breath sounds analysis system. In Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Paris, France, 29 October–1 November 1992; pp. 2582–2583.
23. Hung, K.; Luk, B.L.; Choy, W.H.; Tai, B.; Tso, S.K. Multifunction stethoscope for telemedicine. In Proceedings of the IEEE International Workshop on Computer Architectures for Machine Perception, Hong Kong, China, 26 June–2 July 2004; pp. 87–89.
24. Oletic, D.; Arsenali, B.; Bilas, V. Towards Continuous Wheeze Detection Body Sensor Node as a Core of Asthma Monitoring System. In *Wireless Mobile Communication and Healthcare*; Nikita, K.S., Lin, J.C., Fotiadis, D.I., Waldmeyer, M.-T.A., Eds.; Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Springer: Berlin, Germany, 2012; pp. 165–172.

25. Oletic, D.; Skrapec, M.; Bilas, V. Monitoring Respiratory Sounds: Compressed Sensing Reconstruction via OMP on Android Smartphone. In *Wireless Mobile Communication and Healthcare*; Godara, B., Nikita, K.S., Eds.; Springer: Berlin, Germany, 2013; pp. 114–121.
26. Charleston-Villalobos, S.; Martinez-Hernandez, G.; Gonzalez-Camarena, R.; Chi-Lem, G.; Carrillo, J.G.; Aljama-Corrales, T. Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. *Comput. Biol. Med.* **2011**, *41*, 473–482.
27. Charleston-Villalobos, S.; Albuerne-Sanchez, L.; Gonzalez-Camarena, R.; Mejia-Avila, M.; Carrillo-Rodriguez, G.; Aljama-Corrales, T. Linear and Nonlinear Analysis of Base Lung Sound in Extrinsic Allergic Alveolitis Patients in Comparison to Healthy Subjects. *Methods Inf. Med*. **2004**, *43*, 118–121.
28. Druzgalski, C.K.; Donnerberg, R.L.; Campbell, R.M. Techniques of recording respiratory sounds. *J. Clin. Eng.* **1980**, *5*, 321–330.
29. Kraman, S.S.; Wodicka, G.R.; Pressler, G.A.; Pasterkamp, H. Comparison of lung sound transducers using a bioacoustic transducer testing system. *J. Appl. Physiol*. **2006**, *101*, 469–476.
30. Yadollahi, A.; Moussavi, Z.M.K. A robust method for estimating respiratory flow using tracheal sounds entropy. *IEEE Trans. Biomed. Eng*. **2006**, *53*, 662–668.
31. Yadollahi, A.; Moussavi, Z.M.K. Acoustical Respiratory Flow. *IEEE Eng. Med. Biol. Mag*. **2007**, *26*, 56–61.
32. Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill: New York, NY, USA, 2002.
33. Costa, M.; Goldberger, A.L.; Peng, C.-K. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Phys. Rev. Lett*. **2002**, *89*, 068102.
34. Costa, M.; Goldberger, A.L.; Peng, C.-K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906.
35. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076.
36. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; 2nd ed.; Wiley-Interscience: New York, NY, USA, 2000.
37. Yadollahi, A.; Moussavi, Z.M.K. A robust method for heart sounds localization using lung sounds entropy. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 497–502.
38. Cohen, L. *Time-Frequency Analysis*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1995.
39. Boashash, B. *Time Frequency Analysis*; Gulf Professional Publishing: Oxford, UK, 2003.
40. Gavriely, N.; Palti, Y.; Alroy, G. Spectral characteristics of normal breath sounds. *J. Appl. Physiol*. **1981**, *50*, 307–314.
41. Sierra, G.; Telfort, V.; Popov, B.; Pelletier, M.; Despault, P.; Agarwal, R.; Lanzo, V. Comparison of respiratory rate estimation based on tracheal sounds *versus* a capnograph. In Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society, Shanghai, China, 17–18 January 2006; pp. 6145–6148.
42. Mimoz, O.; Benard, T.; Gaucher, A.; Frasca, D.; Debaene, B. Accuracy of respiratory rate monitoring using a non-invasive acoustic method after general anaesthesia. *Br. J. Anaesth*. **2012**, *108*, 872–875.

43. Que, C.-L.; Kolmaga, C.; Durand, L.-G.; Kelly, S.M.; Macklem, P.T. Phonospirometry for noninvasive measurement of ventilation: methodology and preliminary results. *J. Appl. Physiol*. **2002**, *93*, 1515–1526.
44. Yadollahi, A.; Giannouli, E.; Moussavi, Z. Sleep apnea monitoring and diagnosis based on pulse oximetery and tracheal sound signals. *Med. Biol. Eng. Comput*. **2010**, *48*, 1087–1097.
45. Huq, S.; Moussavi, Z. Acoustic breath-phase detection using tracheal breath sounds. *Med. Biol. Eng. Comput*. **2012**, *50*, 297–308.
46. Al-Khalidi, F.Q.; Saatchi, R.; Burke, D.; Elphick, H.; Tan, S. Respiration rate monitoring methods: A review. *Pediatr. Pulmonol.* **2011**, *46*, 523–529.
47. Ramsay, M.A.E.; Usman, M.; Lagow, E.; Mendoza, M.; Untalan, E.; De Vol, E. The accuracy, precision and reliability of measuring ventilatory rate and detecting ventilatory pause by rainbow acoustic monitoring and capnometry. *Anesth. Analg.* **2013**, *117*, 69–75.
48. Corbishley, P.; Rodriguez-Villegas, E. Breathing Detection: Towards a Miniaturized, Wearable, Battery-Operated Monitoring System. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 196–204.

Variable Frequency Complex Demodulation-Based Motion Artifact Detection for Photoplethysmograph Signals

Duy K. Dao, Jo Woon Chong, S. M. A. Salehizadeh, Ki H. Chon Senior Member, IEEE

***Abstract*—Motion and noise artifacts (MNAs) impose a major limitation on the usability of photoplethysmogram (PPG) especially for ambulatory monitoring. MNAs can distort PPG waveforms causing erroneous estimation of physiological parameters of interest such as heart rate (HR) and oxygen saturation (SpO_2). We present a novel motion and noise artifacts (MNA) detection algorithm based on extraction of time-varying spectral features that are unique to the clean and corrupted components. Using spectral features unique to the noise components of the signal, we derive a noise quality index (*noiseQI*) to quantify and detect MNA corrupted data segments. We examined the sensitivity of the proposed algorithm for MNA detection using clean PPG data corrupted by various levels of additive Gaussian white noise. We utilized the support vector machine (SVM) classifier to build a decision boundary between clean and corrupted data segments from a training data set. We compared our algorithm to three different algorithms known for their alleged prowess in MNA detection: the Hjorth, kurtosis-Shannon Entropy and time-domain variability-SVM approach, which was recently developed in our lab. All these algorithms were tested on PPG data acquired from laboratory-controlled movements with subjects wearing either a forehead or finger pulse oximeter sensor. In addition, we also tested these algorithms on spontaneous PPG data collected from patients enrolled at the University of Massachusetts Emergency Department in Worcester. Our method consistently provided significantly higher detection rate than the other 3 methods, with accuracies greater than 95% for all data. Moreover, our algorithm is able to pinpoint the start and end time of the MNA detection with error less than 1 sec in duration, whereas the next best other algorithm had a detection error more than 2.17 seconds.**

I. INTRODUCTION

Pulse oximeter (PO) is a non-invasive and low cost device which is widely used in clinics to monitor heart rate (HR) and arterial oxygen saturation (SpO_2). Recently, there have been many efforts to derive other physiological parameters from Photoplethysmogram (PPG) recorded by the PO [2]. The fluctuations of PPG contain the influences of arterial, venous, autonomic and respiratory systems of the peripheral circulation. Due to increasing health care costs, a single sensor that has multiple functions such as the PO is very attractive from a financial perspective. Moreover, utilizing a PO as a multipurpose vital sign monitor has a clinical appeal, since the device is widely accepted by clinicians and patients because of its ease of use, comfort and accuracy in providing reliable vital sign. Knowledge of respiratory rate and HR patterns would provide more useful clinical information in many situations in which pulse oximeter is the sole available monitor. However, extraction of the above mentioned vital sign and other physiological parameters using PO is predicated on artifact-free PPG data. It is well known that PPG is highly sensitive to artifacts, particularly those generated while the patient is in motion. This imposes a huge limitation on the usability of PPG for ambulatory monitoring applications. Motion and noise artifacts (MNA) distorting PPG recordings can cause erroneous estimation of HR and SpO2 [3]. While the intelligent design of sensor attachment, form factors and packaging can help to reduce the impact of motion disturbances by making sure that the sensor is securely mounted, they are not sufficient for complete MNA removal. Combating MNA in PPG has been the core focus of research for many years.

Although there are a number of techniques which have been proposed to alleviate the effects of MNA, solution to this problem still remains unsatisfactory in practice. Several algorithm-based MNA reduction methods were proposed, such as time and frequency domain filtering, power spectrum analysis, and blind source separation techniques [4-10]. These techniques reconstruct noise contaminated PPG such that a noise-reduced signal is obtained. However, the reconstructed signal typically contains incomplete dynamic features of the uncorrupted PPG signal and some algorithms are solely designed to capture only the HR and SpO_2 information instead of the signal's morphology and its amplitudes. Moreover, these reconstruction algorithms operate even on clean PPG portions where MNA reduction is not needed. This action introduces unnecessary computation and distorts the signal integrity of the clean portion of the data. Hence, an accurate MNA detection algorithm, which identifies clean PPG recordings from corrupted portions, is essential for the subsequent MNA reduction algorithm so that it does not distort the non-corrupted data segments [11].

MNA detection methods are mostly based on a signal quality index (SQI) which quantifies the severity of the artifacts. Some approaches quantify SQI using waveform morphologies [12-14] or filtered output [15], while other derive SQI with the help of additional hardware such as accelerometer and electrocardiogram [16, 17]. Statistical measures, such as skewness [18], kurtosis, Shannon entropy, and Renyi's entropy, have been shown to be helpful in determining the SQI [19, 20]. These statistical algorithms differentiate the distribution of amplitudes between PPG segments with an assumption that clean and corrupt segments would form two separate groups. However, PPG waveforms vary among patients, thus yielding multitude of amplitude distributions. Therefore, it would be difficult to obtain high accuracy from these algorithms in practice. Our recently published MNA detection method uses time-domain features such as variability in heart rate, amplitude, waveform morphology with the help of the support vector machine (SVM) classifier for detection [21]. The algorithm, which we termed time-domain variability SVM (TDV-

SVM) is shown to be more robust than other statistical-based algorithms as it uses successive difference and variability measures. However, this method is highly dependent on accuracy of the peak amplitude detection. Unlike electrocardiogram (ECG), PPG waveform does not have distinctive peaks which make accurate peak detection challenging. The dependency on a peak detection subroutine is a drawback of the TDV-SVM algorithm and inevitably affects its performance.

Time-frequency (TF) techniques such as Smoothed Pseudo Wigner-Ville, Short Time Fourier Transform, Continuous Wavelet Transform, Hilbert-Huang Transform, and Variable Frequency Complex Demodulation (VFCDM) received considerable attention as means to analysis the signal of interest in both temporal and spectral domains. It is hypothesized in the design of our MNA detection algorithm that TF information would provide meaningful dynamic features for differentiating artifacts. In this paper, we test the efficacy of our proposed algorithm on PPG data sets recorded from the forehead and finger pulse oximeters in laboratory-controlled experiments as well as patient data collected at the University of Massachusetts's Emergency Department. We first introduce time-frequency features derived from VFCDM to quantify the MNA in the recorded PPG signal. The features are then used as inputs for a machine-learning algorithm which utilizes the Support Vector Machine (SVM).

II. Materials and Method

A. Experimental Protocol and Preprocessing

To develop and evaluate our proposed algorithm, we collected PPG data from healthy subjects in lab controlled environment and patients who were admitted to the emergency department at UMASS Hospital. For the laboratory controlled environment, both forehead and finger worn PO sensor data were collected from healthy subjects recruited from the student community of Worcester Polytechnic Institute (WPI). The second data set was acquired from patients who were admitted to our partner hospital at the UMass Memorial Medical Center (UMMC). Laboratory data allows us to have more control over the duration of MNA generated to ensure that the detection algorithms were tested on a wide range of MNA duration. Data from patients provided most realistic information on the motion artifacts since the patients were allowed to move freely as long as the sensors were positioned properly. This study was approved by both WPI's and UMMC's IRBs and all subjects were given informed consents prior to data recordings. PPG data were collected by our custom-made reflectance-type forehead and a transmission-type finger PO.

In *laboratory-controlled head and finger movement* data, motion artifacts were induced by head and finger movements for specific time intervals in both horizontal and vertical directions. For head movement data, 11 healthy volunteers were asked to wear our PO on the forehead along with a reference Masimo Radical (Masimo SET®) finger type transmittance pulse oximeter. After baseline recording for 5 minutes without any movement, subjects were instructed to introduce motion artifacts for specific time intervals varying from 10 to 50% within a 1 minute segment. For example, if a subject was instructed to perform left-right random movements for 6 seconds, an 1 min segment of data would contain 10% noise. The finger laboratory movement data were recorded in a similar setup as the head data using our custom-made PPG finger sensor.

The *patient PPG data* were recorded from 10 subjects admitted to emergence rooms at UMMC. Similar to the laboratory-controlled dataset, each patient was fitted with our custom-made sensors

(both forehead and finger) and the Masimo POs on the forehead and fingers, respectively. The patients were admitted due to pain related symptoms and were not restrained from making natural movements. Therefore, they are expected to generate many different but natural characteristics of MNA in the recorded PPG.

To further evaluate robustness of the proposed algorithm, Gaussian white-noise of 5 minute duration was added to the clean forehead and finger PPG data with varying signal-to-noise ratio (SNR) ranging from 40dB to -5dB. All PPG data were pre-processed by a 6th order infinite impulse response (IIR) band pass filter with cut-off frequencies of 0.1 Hz and 10 Hz. Zero-phase forward and reverse filtering was applied to account for the non-linear phase of the IIR filter.

Many recent publications on MNA detection utilized human visual inspection from experts who were familiar with PPG and their decisions are regard it as the gold standard for marking MNA corrupted data [11, 22, 23]. In our work, we also use the human visual inspection to establish a MNA reference for our datasets. Three inspectors individually marked MNA corrupted portions of the PPG data. Disagreements of the marked portions were resolved by majority votes.

B. VFCDM Features from PPG Signals

VFCDM is a method for estimating time-frequency spectrum (TFS) of a time-varying signal. This method was shown to provide concomitant high time and frequency resolution as well as preservation of the amplitude distribution of the signal [18]. VFCDM has two phases: (1) construct an initial TFS (iTFS) using a method developed in our laboratory, termed fixed frequency complex modulation (FFCDM); (2) the centered frequencies of the iTFS are used for further complex demodulation (CDM) to obtain even more accurate TFS and amplitude of TFS. The VFCDM methodology is detailed as followed.

1) VFCDM

Consider a sinusoidal signal $x(t)$ to be a narrow band oscillation with a time-varying center frequency $f(\tau)$, instantaneous amplitude $A(t)$, phase $\phi(t)$, and the direct current component $dc(t)$:

$$x(t) = dc(t) + A(t)\cos\left(\int_0^t 2\pi f(\tau)d\tau + \phi(t)\right) \tag{1}$$

For a given center frequency, we can extract the instantaneous amplitude information $A(t)$ and phase information $\phi(t)$ by multiplying (3) by $e^{-\int_0^t 2\pi f(\tau)d\tau}$ which results in the following:

$$z(t) = x(t)e^{-j\int_0^t 2\pi f(\tau)d\tau} = dc(t)e^{-j\int_0^t 2\pi f(\tau)d\tau} + \frac{A(t)}{2}e^{j\phi(t)} + \frac{A(t)}{2}e^{-j\left(\int_0^t 2\pi f(\tau)d\tau+\phi(t)\right)} \tag{2}$$

From (2), if $z(t)$ is filtered with an ideal low-pass filter (LPF) with a cutoff frequency$f_c < f_0$, then the filtered signal $z_{lp}(t)$ will contain only the component of interest:

$$z_{lp}(t) = \frac{A(t)}{2}e^{j\phi(t)} \tag{3}$$

The instantaneous frequency is given by $f(t) = f_0 + \frac{1}{2\pi}\frac{d\phi(t)}{dt}$

where f_0 is the centered frequency of interest. By changing the centered frequency followed by using the variable frequency approach as well as the LPF, the signal,$x(t)$, will be decomposed into the sinusoid modulation by the CDM technique as follows:

$$x(t) = \sum_i d_i = dc(t) + \sum_i A_i(t)\cos\left(\int_0^t 2\pi f_i(\tau)d\tau + \phi_i(t)\right) \tag{4}$$

The instantaneous frequency and amplitude of d_i can be calculated using the Hilbert transform

$$X(t) = real(z_{lp}(t)) \qquad Y(t) = imag\left(z_{lp}(t)\right) = H[X(t)] = \frac{1}{\pi}\int\frac{X(t')}{t-t'}dt'$$

$$A(t) = 2\left|z_{lp}(t)\right| = [X^2(t) + Y^2(t)]^{1/2} \quad \phi(t) = \arctan\left(\frac{imag\left(z_{lp}(t)\right)}{real\left(z_{lp}(t)\right)}\right) = \arctan\left(\frac{Y(t)}{X(t)}\right) \tag{5}$$

FFCDM operates by performing CDM on fixed frequency f_0 within confined bandwidth and repeat it over entire frequency band. In order to obtain even higher resolution TFS, center frequencies in iTFS obtained from FFCDM were used for subsequent CDM with finer bandwidth.

2) MNA Discriminative Feature from VFCDM-TFS

An example of VFCDM-TFS is shown in Fig. 1; we term the HR trace and two of its harmonic traces as FM_1, FM_2, FM_3, respectively, with the corresponding amplitudes AM_1, AM_2, AM_3. Our algorithm first determines the dominant frequency in the PPG segment termed f_{HR}. The TFS of the data segment is normalized by the total power in the f_{HR} band. It then extracts AM_1 from a narrow band spectrum of the TFS centered at the dominant frequency $\mathrm{AM}_1 \in [\, f_{HR} - BW, f_{HR} + BW]$. The maximal power in each time instance is taken to form AM_1 in the segment. Once located, AM_1 is removed from the TFS. Similarly, $AM_2 \in [2\mathrm{f_{HR}} - \mathrm{BW}, 2\mathrm{F_{HR}} + \mathrm{BW}]$ and $AM_3 \in [3f_{HR} - BW, 3f_{HR} + BW]$ are found and removed. From the extracted TFS, FMs and AMs, three features were derived to quantify the noise level between clean versus corrupted PPG segments.

a. Residual noise power (P_{noise})

After extracting the first three dominant traces, remaining power in the TFS is considered the residual noise power P_{noise} and is denoted as:

$$P_{noise} = P_{TFS} - \sum_{i=1}^{3} \sum_j AM_{i,j} \tag{6}$$

where P_{noise} is the total power in the TFS. In a clean PPG segment, the first three harmonics would be located within the predetermined narrow band. Thus extracting their power would effectively remove most of the spectral power from the TFS. The remaining noise power would be negligibly small. On the other hand, artifacts in the corrupted PPG segment produce spectral power at various frequency locations which are often not associated with the harmonics' frequency bands. Some of these spectral power would not be extracted which in turn yields high P_{noise} level.

b. Projected frequency modulation difference (df_{FM})

Projected difference is defined as the difference in frequency between the fundamental HR trace and its harmonic traces and is computed as:

$$df_{FM} = \sum_{i=2}^{3} \sum_j FM_{1,j} - i \times FM_{i,j} \tag{7}$$

Similar to the previous assumption, frequency location of the harmonic traces are expected to be proportional to that of the fundamental trace, which would result in a low df_{FM} for a clean segment. For artifact corrupted segment, the proportionality in the frequency of the harmonics would no longer hold, thus driving df_{FM} value to be high.

c. Heart rate frequency difference (df_{HR})

Heart rate frequency difference is defined as the difference between the fundamental frequency modulation FM_1 and HR computed from time-domain peak calculation and is computed as:

$$df_{HR} = FM_1 - \frac{1}{\Delta peak} \tag{8}$$

The three features $P_{noise}, df_{FM}, and\ df_{HR}$ are combined into a weighted sum that represents noise quality index, *noiseQI* in a PPG segment.

$$noiseQI = c_1 P_{noise} + c_2 df_{FM} + c_3 df_{HR} \tag{9}$$

where c_i is the weight of each feature and is determined empirically such that each weighted feature contributes no more than 5% in the clean PPG segments.

It should be noted that the *nosieQI* is very sensitive as noise dynamics contained in only 1 second duration in a 4-second data segment is classified as noise corrupted. This is because the

noise region introduces many frequencies and spectral dynamics that are distinctively dissimilar than the clean signal. Hence, as noted above, P_{noise}, df_{FM}, df_{HR}, and consequently the $noiseQI$ of the MNA corrupted segment will all be significantly larger and distinct from the clean signal's values.

III. SVM-BASED DETECTION OF MOTION/NOISE ARTIFACTS

A. *Feature extraction from PPG*

The preprocessed PPG was used to extract the $nosieQI$, as mentioned in Section II. A sliding window of length L_1 = 8 second with 60% overlap was used to extract the raw PPG signal. The signal was band-pass filtered at 0.1 – 20 Hz and then down-sampled to 20Hz. Only the middle portion of length $L_2 = 4\ second$ of the resulting TFS was considered for further processing. This is because as shown in the Results section, the middle 4 second data length from the initial L_1 = 8 second for the subsequent VFCDM analysis provided the best accuracy in detection of MNA.

To accurate pinpoint time occurrence of MNA, we implemented a *trace-back* strategy, which is triggered when the *noiseQI* value changes its state as illustrated in Fig 2. When *noiseQI* goes from lower than a threshold value of 0.2 to greater than 0.2, the trace-back algorithm computes a new *noiseQI* three times with shifting backward a second at each time instant. For example, in Fig. 2A *noiseQI* changes to a value that is greater than 0.2 at time duration 4-8 seconds. The trace-back scheme would call VFCDM routine to compute new *noiseQI* values for the back-shifted segments at time durations starting at 3-7 seconds, 2-6 seconds, and ending at 1-5 seconds. The threshold $noiseQI => 0.2$ was empirically determined to represent the presence of MNA. As detailed above, our VFCDM algorithm is designed to indicate that a segment is corrupted even if only 1 second of the 4 second duration data contain MNA. Hence, since the 3-7 second segment is determined to be corrupted, it allows us to deduce that the 8th second time point is corrupted. The same logic applies to the 2-6 and 1-5 second segments. If MNA starts at the 5th second, it is expected that the updated *nosieQI* values would be low prior to the 5th second while remain high for the segments after the 5th second. Similarly, MNA end point is determined by applying trace-back algorithm when *noiseQI* goes from higher than 0.2 to lower than 0.2. An example of the track-back strategy on an actual PPG signal is illustrated in Fig. 3.

B. *Classification by Support Vector Machine (SVM)*

SVM was applied to build a decision boundary to classify the MNA segment from the clean PPG data. SVM is widely used for classification and regression analysis due to its accuracy and robustness to noise [25]. The SVM consists of training and test phases as briefly described in the following sections.

The SVM takes *a priori* determined classification parameter values of the clean and corrupted PPG segments as a training data set, finds the support vectors among the training data set which maximize the margin (or the distance) between different classes, and then builds a decision boundary. If the estimated decision is different from its known label, the decision is regarded as a *training error*. We consider a soft-margin SVM which can set the boundary even when the data sets are mixed and cannot be separated. In the soft-margin SVM algorithm, slack variables are introduced to minimize the training error while maximizing the margin. Soft-margin SVM uses the following equation to find the support vectors [26].

$$\text{Minimize } C\sum_{sv=1}^{N}\delta_{sv}+\frac{1}{2}\langle \mathbf{w}_s,\mathbf{w}_s\rangle,$$

$$\text{Subject to } T_{sv}\left(\langle \mathbf{w}_s,\mathbf{y}_{sv}\rangle+b_s\right)\geq 1=\delta_{sv} \text{ for } sv=1,2,\ldots,N \text{ =1,2,…,N, and } \delta_{sv}\geq 0 \tag{19}$$

where C is a regulation parameter, N is the number of vectors, δ_{sv} is the slack variable, $\mathbf{w}_s$ is weight vector and $<\cdot,\cdot>$ is the inner product operation. The T_{sv} is the sv^{th} target variable, $\mathbf{y}_{sv}$ is the sv^{th} input vector data, and b_s is the bias. The SVM decision boundary F_{sv} is derived as

$$F_{sv}=\left\langle \mathbf{w}_s^*,\mathbf{y}\right\rangle+b_s^*=0 \tag{20}$$

where $\mathbf{w}_s^*$ and b_s^* are weight factor and bias, respectively, obtained from Eq. (20) and $\mathbf{y}$ is the input point. By transforming the $\mathbf{y}_{sv}$ and $\mathbf{y}$ terms to $\mathbf{y}_{sv}\rightarrow\Phi(\mathbf{y}_{sv})$ and $\mathbf{y}\rightarrow\Phi(\mathbf{y})$, the non-linear SVM can be transformed to a linear SVM. For nonlinear SVM, Eq. (20) is modified as

$$T_{sv}\left(\left\langle \mathbf{w}_s,\Phi\left(\mathbf{y}_{sv}\right)\right\rangle+b_s\right)\geq 1 \tag{21}$$

To facilitate the operation in nonlinear SVM, a kernel function $K_s(\cdot,\cdot)$, which is a dot-product in the transformed feature space is used as the following:

$$K_s\left(\mathbf{y}_{sv},\mathbf{y}_{sv'}\right)=\left\langle \Phi\left(\mathbf{y}_{sv}\right),\Phi\left(\mathbf{y}_{sv'}\right)\right\rangle \tag{22}$$

where $sv'=1,2,...,N$. Once training phase is completed, the optimal support vectors for training data are determined. These vectors are then used to classify the testing data so that the optimal solution of mutually exclusive boundaries can be determined.

IV. Results

We evaluated the performance of the VFCDM-based MNA detection algorithm for various types (laboratory controlled, and real patients at hospital) of motion-corrupted PPGs to validate its performance for a wide range of scenarios. K-fold cross validation was adopted to evaluate the performance of our algorithm. Specifically, for a dataset of N subjects, data from N-1 subject was used for training and the unused subject data were used for testing. The train-test cycle was done N times, each time with a different test subject. We optimized the regularization parameter value (C = 10) for the linear kernel SVM by minimizing the training error. Fig. 4 shows a representative illustration of the performance of our proposed approach. Fig. 4A displays the filtered PPG signal with contains both clean and MNA which is corrupted with varying levels of amplitude fluctuations. Fig. 4B depicts the corresponding *noiseQI* which was obtained according to Eq. (9). As shown, the *noiseQI* has low values for the clean portion of the data whereas it is high where the MNA occurs. The *noiseQI* is used for the SVM classifier to determine whether the given segment is clean or corrupted. Fig. 4C shows the classification results via our approach along with the trained experts' decision of the occurrence of the MNA for comparison. For this example, we observe that our approach has a high sensitivity and specificity in detecting MNA.

The optimal window length was determined by varying L_2 from 3 to 6 sec while keeping L_1 constant at 8 sec. Detection performance was evaluated by comparing our classification results to the MNA reference (as determined by the experts) to yield accuracy, sensitivity, and specificity. Table I shows performance statistics in terms of accuracy (Acc), sensitivity (Sen), and specificity (Spe) of our MNA detection algorithm at various window length (L_2) for the laboratory dataset. The window length of 4 sec ($L_2 = 4$ second) yielded the best performance in term of accuracy among the various window lengths examined.

The *noiseQI* value indicates the amount of noise present in the PPG segment. It is of interest to evaluate how well this parameter correlates with the actual noise at various levels. Therefore, the nosieQI values were computed for simulated data consisting of clean forehead and finger PPG data corrupted by additive white noise of varying SNR levels. The results are shown in Fig.5. As shown, the *noiseQI* is consistently low for signal with high SNR values and high for signal with low SNR values with reaching a unity value for SNR less than 0 dB for both finger and forehead data.

We compared the proposed algorithm with three other recently published MNA detection algorithms: 1) Hjorth features (Hjorth), 2) time-domain variability SVM (TDV-SVM) approach, 3) Kurtosis-Shannon Entropy (KSE) features [11, 22, 23, 27]. A representative example of the MNA detection comparing all of the aforementioned methods is illustrated in Fig. 6. Performance results of each method are presented in Table II. Two-sample one-sided Welch's t-tests at 95% confident interval was performed to assess the significance of the accuracy, sensitivity, and specificity results from the K-fold cross-validation. In order to measure and compare the detection powers, receiver operative characteristic curves were generated for all the features used in our VFCDM algorithm and other detection algorithms. Areas under these curves (AUCs) represent the strength of these features. In Fig. 7, ROC curves with highest AUC values were shown. Note that our VFCDM-derived features obtained the highest AUC values among all methods compared.

In addition to accurate MNA detection, the other attractive feature of our algorithm is that it is able to accurately locate the start and end points of MNA occurrences. To evaluate the algorithm's effectiveness in pinpointing the start and end time of the MNA, we computed the time difference of start and end points between the visual reference and detection algorithms' results. The time difference is termed detection transition time, *DTT*, which reflects how accurate on average a MNA algorithm detects the start and end time of the MNA. Table III provides *DTT* comparison of our VFCDM algorithm and other detection algorithms. As shown in Table III, our algorithm's detection accuracy of the duration of the MNA is significantly better than 3 other methods. Our algorithm is off by less than 1 second whereas the second best algorithm, the Hjorth, is more than 2 seconds off and the least accurate method, the KSE, is off by more than 4 seconds.

V. Discussion

We propose an accurate MNA detection method which uses dynamic characteristics of the corrupted PPG derived via the VFCDM. The efficacy of the detection method was validated using contrived motion data from healthy subjects and unconstrained MNA data from patients at UMMS ER department. In this study, several key features associated with MNA, derived from the VFCDM-based time-frequency spectrum, were utilized for detection of MNA. By transforming the PPG times series into time-frequency domain, we were able to better capture time-varying characteristics of the MNA. Specifically, we recognized that PPG's clean signal dynamics are largely concentrated at the heart rate and its harmonic frequency bands, hence, we surmise that the presence of large amplitudes in the other frequency bands must be associated with MNA. This is clearly seen in Fig. 1B as VFCDM results from a clean PPG signal yields distinct peaks across all times at the HR frequencies and its 2 successive harmonic frequencies. Therefore, we divided the TFS into three narrow band spectrums and tracked down these amplitude traces accordingly. In a clean PPG segment, Fig. 1A-B, most of the spectral power is concentrated in the $FM_1, FM_2,$ and FM_3 traces since the signal is sinusoidal-like and periodic in nature. In a MNA corrupted PPG segment shown in Fig. 1C-D, however, the signal is disturbed by inconsistent changes in the signal amplitude due to motion. These changes are typically irregular thus creating various spectral contents in the resulting TFS and eventually yielding high *noiseQI* values as

defined in Eq. (8). Hence, the *noiseQI* values provide a quantitative measure of the MNA that is present in the PPG signal.

Results in Fig. 5 indicate that if we set the SNR cutoff threshold at 15dB to discriminate between the clean and corrupted PPG data segment, then the corresponding *noiseQI* cutoff threshold would be around 0.14, and 0.22 for the head and finger data, respectively. Finger PPG waveforms for some people contain prominent traces of dicrotic notch, which results in additional high frequency components. Our VFCDM algorithm assumes that the heart rate cycle is the dominant component of the PPG waveform and the remaining components are considered as noise. Therefore, the *noiseQI* values of the finger PPG are higher than of the head PPG due to additional high frequency components introduced by the dicrotic notch. It can also be speculated that PPG recorded from peripheral part (finger) is modulated at a higher intensity than that of forehead, which may explain higher *noiseQI* values for the former location. Hence, due to the intrinsic difference in morphological properties of PPG recorded from the finger and forehead, training for MNA detection needs to be done separately for these two measurement locations.

The detection accuracy on both lab-controlled and UMMC datasets using our proposed methods outperformed the other three detection methods: Hjorth, TDV-SVM, and KSE. We first compared each method's detectability based on their own unique feature selection by evaluating the area under ROC curve. The AUC showed that our *noiseQI* feature provided that highest value at 0.986 for both finger and forehead PPG, as shown in Fig. 7. Concomitantly, the accuracy, sensitivity and specificity values of our method were significantly higher than all other methods as indicated in Table II.

The eventual aim of our proposed algorithm is to detect MNA in real time. The algorithm only takes 33.3ms to compute *noiseQI* for a 4s PPG window length using Matlab on a PC with Intel Xeon at 3.6GHz. Therefore, it would be straightforward to optimize the algorithm for real time detection of MNA in PPG signal.

In conclusion, we proposed an accurate MNA detection algorithm that utilizes both time and spectral features to classify between a clean and corrupted PPG data segment. Comparison of our algorithm using both lab-controlled and clinical data to three other detection methods showed far better performance of our algorithm. Moreover, it was found that our algorithm was superior to other algorithms in detecting the start and end time of the MNA. Accurate detection of the start and end time of the MNA is important for the subsequent reconstruction of the MNA-corrupted data as we do not want to miss the MNA portion of data for reconstruction or avoid having to reconstruct when the data segment is designated to be clean. Moreover, our algorithm is real-time realizable and it is applicable to either transmission (finger) or reflectance (forehead) PPG sensor.

TABLE I. MEAN ± STD. DEVIATION OF PERFORMANCE METRICS OF OUR PROPOSED VFCDM USING VARIOUS WINDOW LENGTH

L		3s	4s	5s	6s
Lab. Head	Acc	86.9 ± 5.54	91.9 ± 3.11	87.2 ± 4.23	84.1 ± 9.26
	Sen	81.8 ± 9.69	88.8 ± 5.33	80.3 ± 6.78	75.5 ± 16.40
	Spe	89.4 ± 6.17	93.9 ± 2.69	80.3 ± 3.11	89.3 ± 7.95

TABLE II. MEAN ± STD. DEVIATION OF PERFORMANCE METRICS OF OUR PROPOSED VFCDM, OTHER METHODS . (*) INDICATE STATISTICAL SIGNIFICANCE (P<0.01) BETWEEN OUR METHOD VERSUS THE OTHERS.

		VFCDM	Hjorth		TDV		KSE	
Lab. Head	Acc	96.9 ± 1.4	72.5 ± 10.7	*	93.4 ± 3.5	*	83.1 ± 7.3	*
	Sen	96.3 ± 2.4	47.2 ± 28.8	*	88.8 ± 7.9	*	56.6 ± 17.9	*
	Spe	97.7 ± 2.0	84.4 ± 4.6	*	96.7 ± 3.0	*	91.5 ± 1.1	*
Umass. Head	Acc	95.5 ± 2.2	69.5 ± 21.8	*	89.8 ± 1.4	*	85.1 ± 8.3	*
	Sen	91.1 ± 3.5	53.8 ± 26.1	*	84.6 ± 2.9	*	68.7 ± 17.7	*
	Spe	96.3 ± 1.4	84.5 ± 8.8	*	94.3 ± 4.0		86.4 ± 7.8	*
Lab. Finger	Acc	97.7 ± 1.4	91.1 ± 6.9	*	94.4 ± 3.3	*	58.5 ± 20.7	*
	Sen	96.3 ± 2.6	83.5 ± 21.0	*	94.7 ± 3.4	*	34.6 ± 12.2	*
	Spe	98.2 ± 1.4	96.2 ± 3.0		94.7 ± 3.0	*	86.3 ± 15.7	*
Umass. Finger	Acc	95.5 ± 2.4	71.0 ± 19.0	*	89.6 ± 2.3	*	88.3 ± 2.7	*
	Sen	89.2 ± 2.0	41.1 ± 27.6	*	85.2 ± 3.8	*	71.5 ± 8.8	*
	Spe	97.2 ± 2.1	88.4 ± 7.5	*	95.1 ± 0.8		93.6 ± 1.5	*

TABLE III. MEAN ± STD. DETECTION OF TRANSITION TIME, DTT OF OUR VFCDM METHOD AND OTHER METHODS. (*) INDICATE STATISTICAL SIGNIFICANCE (P<0.01) BETWEEN OUR METHOD VERSUS THE OTHERS.

Algorithm	DTT (sec)	
VFCDM	0.94 ± 0.65	
Hjorth	2.17 ± 0.37	*
KSE	4.24 ± 2.42	*
TDV	2.75 ± 0.96	*

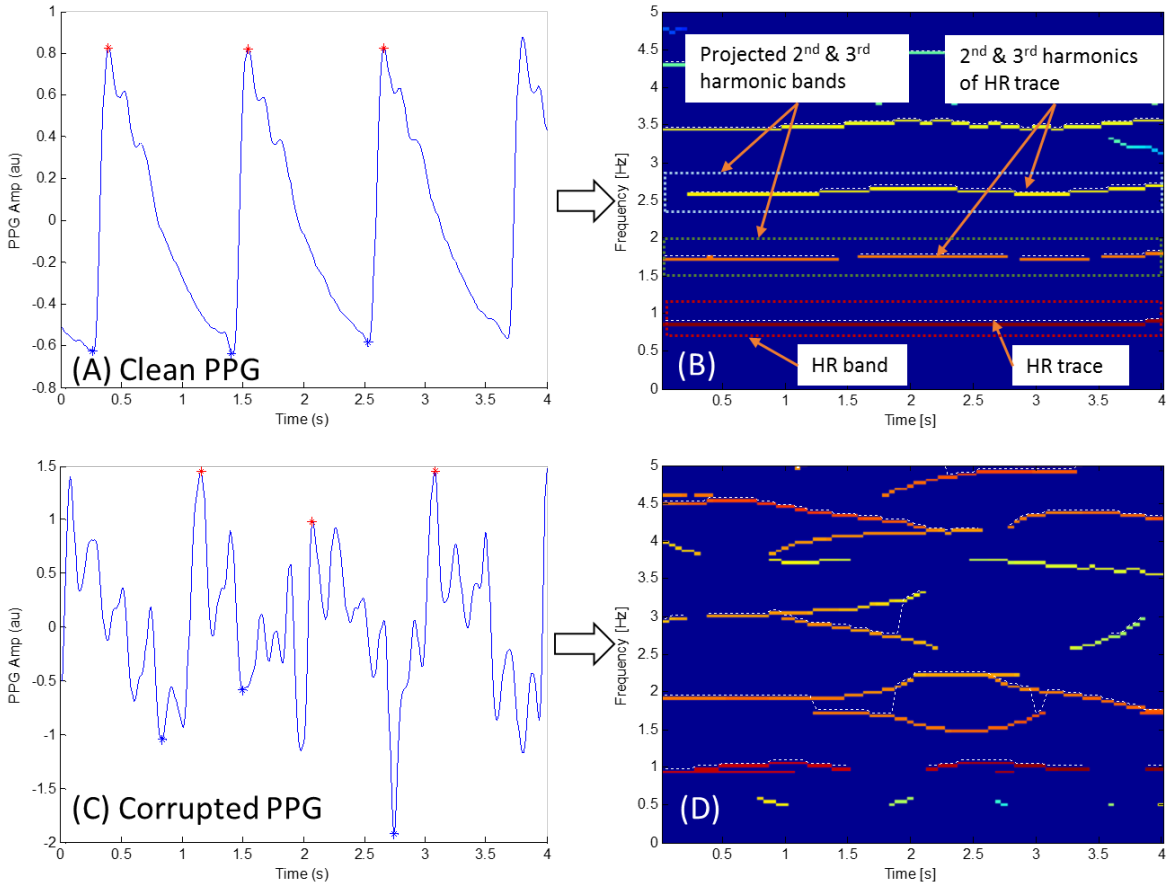


Figure 1. Time-Frequency Spectrum produced by VFCDM in 8sec PPG window (L = 4s). The rectangular bands in red, yellow, and green are HR band and its 2nd and 3rd projected harmonic bands. The white dotted lines in their corresponding bands are extracted HR and its harmonic traces. (A) Clean PPG. (B) TFS of clean PPG. (C) Corrupted PPG. (D) TFS of corrupted PPG.

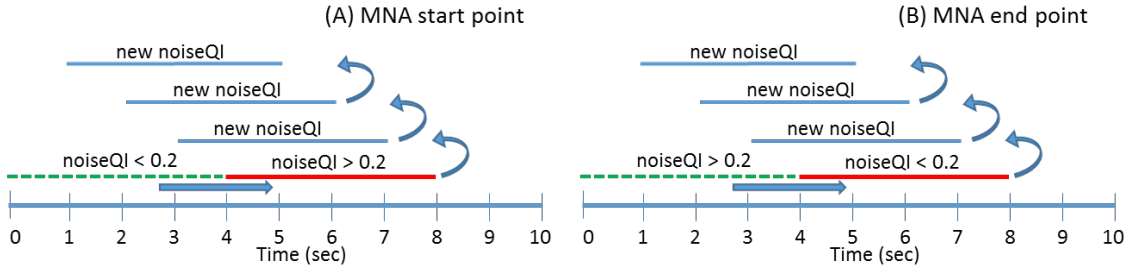


Figure 2. Trace-back strategy to determine the start and end points of MNA.

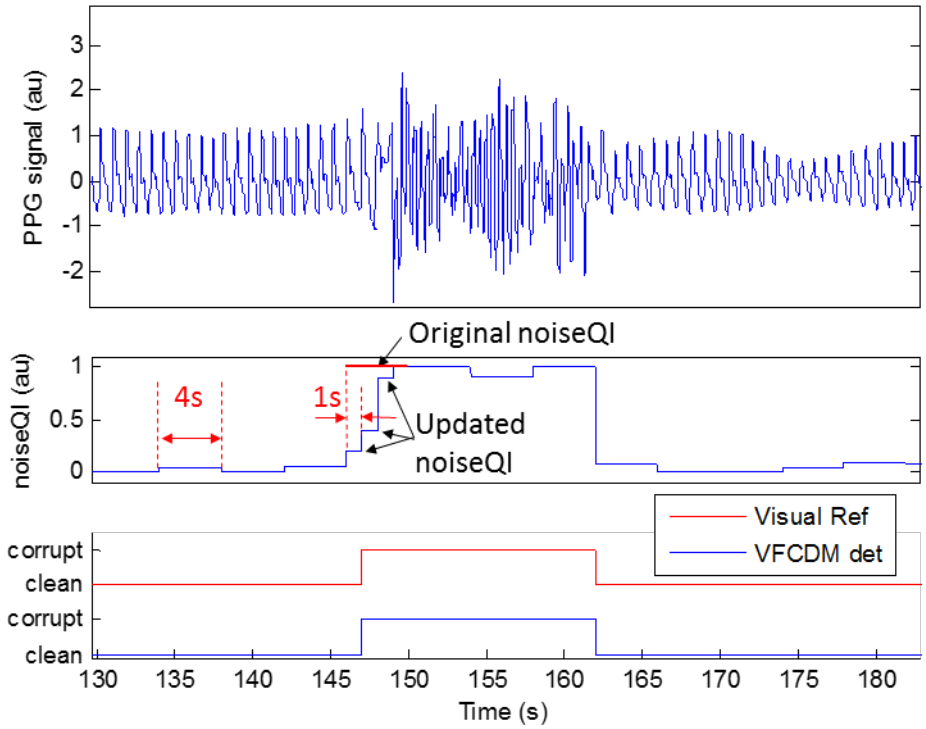


Figure 3. Illustrative example of the trace-back mechanism. In normal mode, our algorithm compute *noiseQI* at every 4 second interval. When triggered, trace-back mechanism update nosieQI very 1 sec backward, which helps accurately determine the start and end point of MNA.

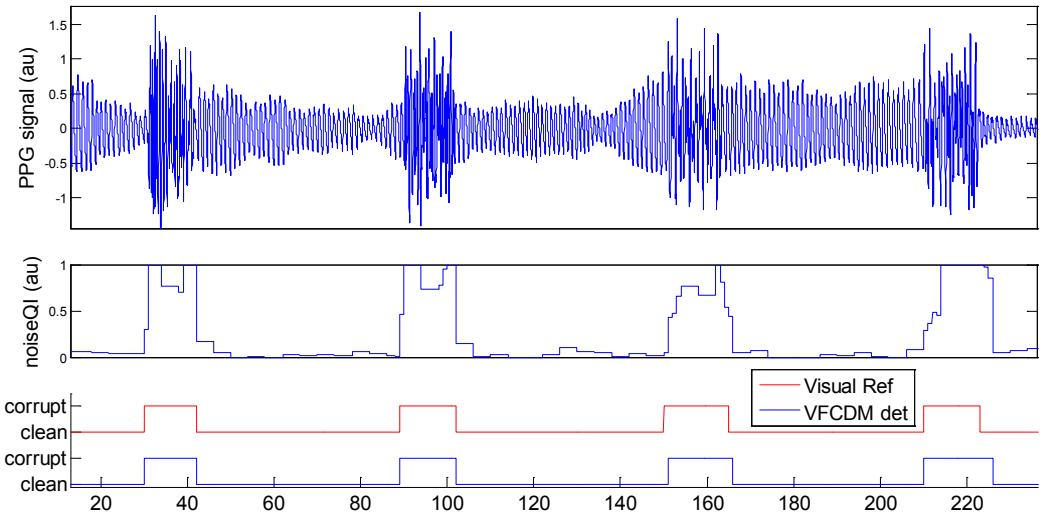


Figure 4. An example of MNA detection using VFCDM. (A) PPG signal corrupted with MNA. (B) *noiseQI* derived from VFCDM. (C) Detection decision of VFCDM (blue) and reference MNA (red)

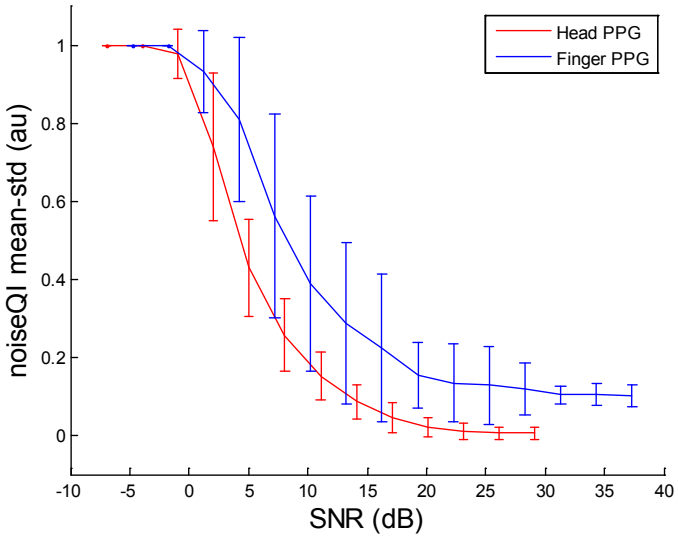


Figure 5. Noise quality index (*noiseQI*) computed from the additive white-noise-corrupted PPG data at various signal-to-noise ratios (SNR). Red: PPG recorded from forehead; Blue: PPG recorded from finger.

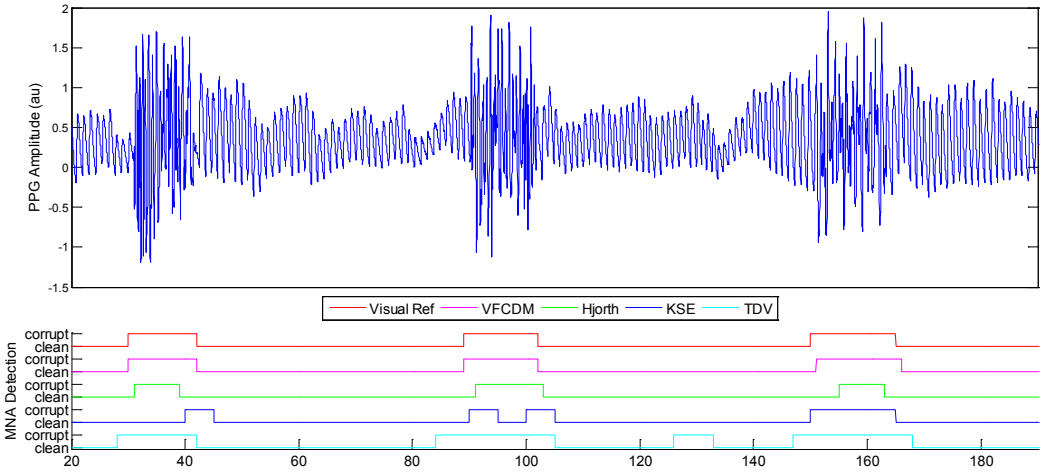


Figure 6. An example of MNA detection using our VFCDM method versus other methods: Hjorth, Time-domain and Kurtosis-Shannon Entropy. The pulse-like traces are the MNA reference and detection results from the feature sets. High value indicates detected MNA, otherwise clean PPG signal.

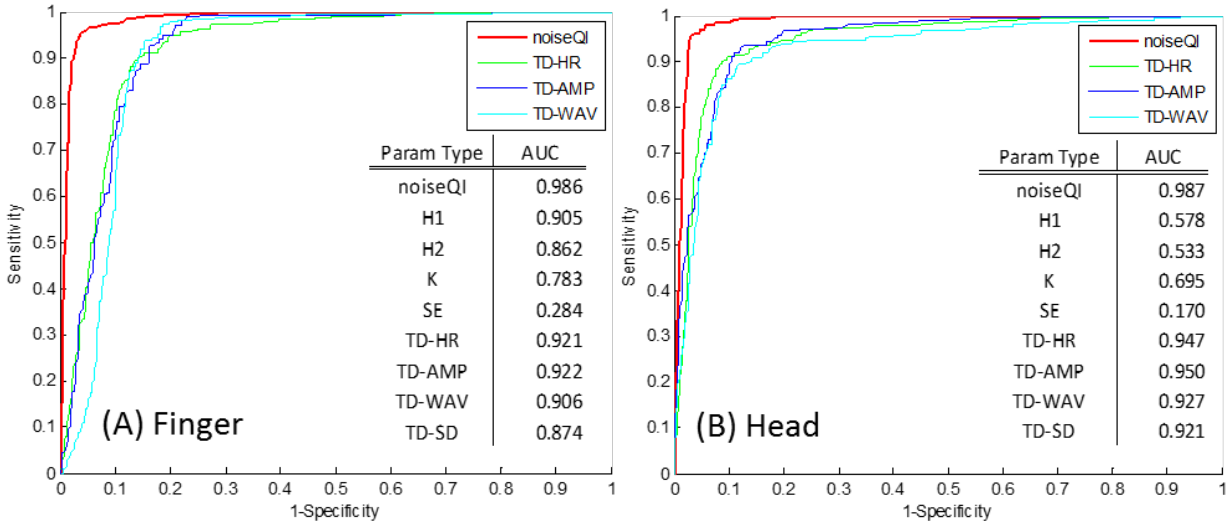


Figure 7. Receiver-operative-curves (ROCs) of all the features used in MNA detection algorithms: Our proposed VFCDM (*noiseQI*); Hjorth parameters (H1,H2); Statistical features (K,SE), Time-domain features (TD-HR, TD-AMP,TD-WAV,TD-SD). (A) Finger data; (B) Forehead data.

REFERENCES

[1] M. T. Petterson, V. L. Begnoche, and J. M. Graybeal, "The effect of motion on pulse oximetry and its clinical significance," *Anesth Analg,* vol. 105, pp. S78-84, Dec 2007.

[2] K. Nakajima, T. Tamura, and H. Miike, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique," *Medical Engineering & Physics,* vol. 18, pp. 365-372, 7// 1996.

[3] R. Sahni, A. Gupta, K. Ohira-Kist, and T. S. Rosen, "Motion resistant pulse oximetry in neonates," *Arch Dis Child Fetal Neonatal Ed,* vol. 88, pp. F505-8, Nov 2003.

[4] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A Novel Approach for Motion Artifact Reduction in PPG Signals Based on AS-LMS Adaptive Filter," *Instrumentation and Measurement, IEEE Transactions on,* vol. 61, pp. 1445-1457, 2012.

[5] T. L. Rusch, R. Sankar, and J. E. Scharf, "Signal processing methods for pulse oximetry," *Comput Biol Med,* vol. 26, pp. 143-59, Mar 1996.

[6] R. Hong Enriquez, M. Sautie Castellanos, J. Falcon Rodriguez, and J. L. Hernandez Caceres, "Analysis of the photoplethysmographic signal by means of the decomposition in principal components," *Physiol Meas,* vol. 23, pp. N17-29, Aug 2002.

[7] B. S. Kim and S. K. Yoo, "Motion artifact reduction in photoplethysmography using independent component analysis," *Biomedical Engineering, IEEE Transactions on,* vol. 53, pp. 566-568, 2006.

[8] Y. S. Yan, C. C. Poon, and Y. T. Zhang, "Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner-Ville distribution," *J Neuroeng Rehabil,* vol. 2, p. 3, Mar 1 2005.

[9] J. Lee, W. Jung, I. Kang, Y. Kim, and G. Lee, "Design of filter to reject motion artifact of pulse oximetry," *Computer Standards & Interfaces,* vol. 26, pp. 241-249, 2004.

[10] M. R. Ram, K. V. Madhav, E. H. Krishna, K. N. Reddy, and K. A. Reddy, "Use of spectral estimation methods for computation of SpO_2 from artifact reduced PPG signals," in *Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE*, 2011, pp. 431-436.

[11] R. Krishnan, B. Natarajan, and S. Warren, "Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data," *Biomedical Engineering, IEEE Transactions on,* vol. 57, pp. 1867-1876, 2010.

[12] J. A. Sukor, S. J. Redmond, and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiol Meas,* vol. 32, pp. 369-84, Mar 2011.

[13] K. Li, S. Warren, and B. Natarajan, "Onboard Tagging for Real-Time Quality Assessment of Photoplethysmograms Acquired by a Wireless Reflectance Pulse Oximeter," *Biomedical Circuits and Systems, IEEE Transactions on,* vol. 6, pp. 54-63, 2012.

[14] 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, pp. 15-32, Jan 2008.

[15] W. Karlen, K. Kobayashi, J. M. Ansermino, and G. A. Dumont, "Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation," *Physiol Meas,* vol. 33, pp. 1617-29, Oct 2012.

[16] B. Lee, J. Han, H. J. Baek, J. H. Shin, K. S. Park, and W. J. Yi, "Improved elimination of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry," *Physiol Meas,* vol. 31, pp. 1585-603, Dec 2010.

[17] J. Y. Foo and S. J. Wilson, "A computational system to optimise noise rejection in photoplethysmography signals during motion or poor perfusion states," *Med Biol Eng Comput,* vol. 44, pp. 140-5, Mar 2006.

[18] H. Wang, K. Siu, K. Ju, and K. H. Chon, "A high resolution approach to estimating time-frequency spectra and their amplitudes," *Ann Biomed Eng,* vol. 34, pp. 326-38, Feb 2006.

[19] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, "Automatic real time detection of atrial fibrillation," *Ann Biomed Eng,* vol. 37, pp. 1701-9, Sep 2009.

[20] C. Yu, Z. Liu, T. McKenna, A. T. Reisner, and J. Reifman, "A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms," *J Am Med Inform Assoc,* vol. 13, pp. 309-20, May-Jun 2006.
[21] S. M. Salehizadeh, D. K. Dao, J. W. Chong, D. McManus, C. Darling, Y. Mendelson, *et al.*, "Photoplethysmograph Signal Reconstruction based on a Novel Motion Artifact Detection-Reduction Approach. Part II: Motion and Noise Artifact Removal," *Ann Biomed Eng,* May 14 2014.
[22] J. W. Chong, D. K. Dao, S. M. Salehizadeh, D. D. McManus, C. E. Darling, K. H. Chon, *et al.*, "Photoplethysmograph Signal Reconstruction Based on a Novel Hybrid Motion Artifact Detection-Reduction Approach. Part I: Motion and Noise Artifact Detection," *Ann Biomed Eng,* Aug 5 2014.
[23] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 4972-4975.
[24] R. Zou, H. Wang, and K. H. Chon, "A robust time-varying identification algorithm using basis functions," *Ann Biomed Eng,* vol. 31, pp. 840-53, Jul-Aug 2003.
[25] K. V. P. Naraharisetti, M. Bawa, and M. Tahernezhadi, "Comparison of different signal processing methods for reducing artifacts from photoplethysmograph signal," in *Electro/Information Technology (EIT), 2011 IEEE International Conference on*, 2011, pp. 1-8.
[26] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," Department of Computer Science, National Taiwan University2003.
[27] B. Hjorth, "The physical significance of time domain descriptors in EEG analysis," *Electroencephalography and Clinical Neurophysiology,* vol. 34, pp. 321-325, 1973.