

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 2015

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE October 2015		2. REPORT TYPE Annual		3. DATES COVERED 25 Sep 2014 - 24 Sep 2015	
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: ki.chon@uconn.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	
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 Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			USAMRMC
			Unclassified		19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
1. Introduction.....	4
2. Keywords.....	4
3. Accomplishments.....	4
4. Impact.....	5
5. Changes/Problems.....	6
6. Products.....	6
7. Participants & Other Collaborating Organizations.....	7
8. Special Reporting Requirements.....	7
9. Appendices.....	7

1. 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.

2. Keywords

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

3. Key Research Accomplishments

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.

Accomplishments:

Aim1

- The Aim 1 was completed in the year 2 and the following was reported in the annual report 2:
- 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 methods. For all new data including that from UMASS, our new MNA reconstruction algorithm consistently provide better accuracy than our previously-published algorithms. A manuscript describing this new algorithm has been accepted for publication.

- We developed a new algorithm for accurate reconstruction of heart rates for the MNA-corrupted portions of the data. This manuscript is accepted for publication. This algorithm is more computationally efficient than our recently-published algorithm. Moreover, our algorithm provides the best performance when compared to other published algorithms using the same datasets (IEEE Signal Processing Cup Challenge datasets).
- We have signed a formal agreement between UConn (Dr. Chon's lab) and Samsung to use their smartwatch, called Simband, for implementing our atrial fibrillation detection and motion artifact detection/reconstruction algorithms directly onto their smartwatch. We were chosen as a developer for this development with Samsung. The Simband is not yet released to the public. The Simband hardware contains PPG and ECG sensors. Moreover, it has sensors for temperature and impedance measurements. We are excited about this opportunity as we can fully test our motion artifact algorithms we have developed during the current Army funding for the next 6 months (2016). We will be conducting a clinical study at UMass and will be distributing these Simbands to patients to collect PPG data for AF detection. These subjects will be wearing a Holter monitor and Simband simultaneously for 2 weeks. This data will allow us to examine the effectiveness of our MNA algorithms in a real-life data acquisition setting.
- We received a 4-year grant from the National Science Foundation (NSF) in the amount of \$1.15M to develop a wearable vest and a smartwatch for early detection of heart failure and atrial fibrillation. We will incorporate our motion artifact detection and reconstruction algorithms in a smartwatch for atrial fibrillation detection. We are grateful for the Army's support which enabled us to develop a robust algorithms for MNA detection and reconstruction of heart rates; this technology was instrumental in obtaining a grant from the NSF. The PI is Dr. Ki Chon and Co-PIs are Drs. Darling and McManus from UMASS Medical School and Dr. Mendelson from WPI. The grant commenced this Fall (09/01/15-08/31/19).
- We have published 7 journal articles and 3 journal articles will be submitted in the coming month based on algorithm development.
- We will file three new patent disclosures on our algorithms.
- We developed a new and simple method to calibrate and estimate tidal volume from a smartphone without using any external sensors; a patent disclosure will be 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 137 patients enrolled in our study.
- Blood loss characterization using our algorithm, which detects the reduction of amplitude modulation values in the heart rate frequency range of the time-varying spectrum of PPG data, was applied to 24 trauma and 27 dehydration subjects' data collected at UMass Med. Our algorithm continues to provide promising results as we have found 92% accuracy, 100% sensitivity and 89.5% specificity on detection of blood loss on 24 trauma subjects. The results are also excellent for 27 dehydration subjects, as we found 85% accuracy, 100% sensitivity and 79% specificity.

Opportunities for training and professional development:

- Nothing to report.

How were results disseminated to communities of interest?

- Our results have been disseminated via journal publications and conference proceedings. We have also attended several conferences to disseminate our results.

What do you plan to do during the next reporting period?

- We will continue to enroll trauma and dehydration subjects at UMASS and complete data analysis on these data to examine if our algorithm is able to accurately detect hypovolemia.

4. Impact

What was the impact on the development of the principal discipline(s) of the project?

- Our motion artifact detection and reconstruction algorithms have the best accuracies when benchmarked against other published algorithms. This benchmark comparison of our algorithms against other published methods was made possible using the IEEE Signal Processing Cup Challenge databases.

What was the impact on other disciplines?

- Dr. Chon's lab has signed an official agreement with Samsung Corporation to become a developer on their smartwatch, called Simband, for development and testing of MNA algorithms. This is exciting because, as a developer, we are able to embed our MNA algorithms directly onto the smartwatch and test their efficacy on subjects with atrial fibrillation. The subjects with atrial fibrillation will wear the Simband continuously for two weeks. This opportunity will allow us to fully test our MNA algorithms in a real-life setting.

What was the impact on technology transfer?

- Dr. Chon has created a new start-up company and recruited Mr. Bryant Guffey as the CEO. The company will develop a smartwatch for continuous monitoring of atrial fibrillation. Our motion artifact detection and reconstruction algorithms will be embedded onto a microprocessor in the smartwatch for real-time detection and reconstruction of heart rates and oxygen saturation values for those data segments corrupted with motion artifacts.

What was the impact on society beyond science and technology?

- Tissue and regenerative engineering is the most popular subject among biomedical engineering (BME) students. However, bioinstrumentation, devices, sensors and signal processing topics are not as popular among BME students. The PI has been able to show the progress on the development of smartwatches to BME students during his senior design course lectures and seminars at UConn. These presentations have resulted in significant interest among BME students to focus on their undergraduate studies in the topics of devices, sensors and signal processing.

5. Changes/Problems:

- Nothing to report.

6. Products:

Publications, Abstracts, and Presentations

Refereed Journal Articles

1. Nam, Y., B. Reyes, and **K.H. Chon**, Estimation of respiratory rates using built-in microphone and ear microphone of a smartphone, *IEEE J. Biomedical and Health Informatics*, In Press.
2. Salehizadeh, S.M.A., D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, C. Darling and **K.H. Chon**. A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intensive physical activities using a wearable photoplethysmogram sensor, *Sensors*, In Press.
3. Lazaro, J., Y. Nam, E. Gil, P. Laguna, and **K.H. Chon**, Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals, *Physiological Measurements*, 36:2317-2333, 2015.
4. Reljin, N., B.A. Reyes, and **K.H. Chon**, Tidal volume estimation using blanket fractal dimension of the tracheal sounds acquired by smartphone, *Sensors*, 15:9773-90, 2015.
5. 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, Accepted with revision, *IEEE J. Biomedical and Health Informatics*.
6. Reyes, B.A., N. Reljin, Y. Kong, Y. Nam and **K.H. Chon**, Towards the development of a mobile phonopneumogram: automatic breath-phase classification using smartphones, Accepted with revision, *Annals of BME*.
7. Reyes, B.A., N. Reljin, Y. Kong, Y. Nam, S.H. Ha and **K.H. Chon**, Tidal volume and instantaneous respiratory rate estimation using smartphone camera, Accepted with revision, *IEEE J. Biomedical and Health Informatics*.

Conference Proceedings:

1. Syed Mohamed Amin Salehizadeh, Duy Dao, Yitzhak Mendelson and Chon, K.H., Heart rate monitoring during intense physical activity using PPG signal frequency component analysis, *MHSRS*, Ft. Lauderdale, FL, 2015.
2. Syed Mohamed Amin Salehizadeh, Duy Dao, Yitzhak Mendelson and Chon, K.H., Heart rate monitoring during intense physical activity using photoplethysmogram signal frequency component analysis, *Body Sensor Networks*, Boston, MA, 2015.
3. Hugo F. Posada-Quintero, B.A. Reyes, S.A. Amir, P. Vardakas, H. DiSpirito, K. Burnham, J. Pennace, and Chon, K.H., Developing pressure sensitive adhesive electrodes: preliminary results, *Conf Proc IEEE Eng Med Biol Soc*, 2014, Chicago, IL, 2014: 2742-4.
4. Bersain Reyes, H.F. Posada-Quintero, J.R. Bales and Chon, K.H., Performance evaluation of carbon black based electrodes for underwater ECG monitoring, *Conf Proc IEEE Eng Med Biol Soc*, 2014, Chicago, IL, 2014:1691-4.

Inventions, Patents and Licenses

1. Calibration of Tidal Volume Using Video Camera Images of Smartphones, Invention disclosure filed with UConn. Inventors: Ki Chon and Bersain Reyes

2. Motion and Noise Artifact Detection and Reconstruction Algorithms for Photoplethysmogram Signals: estimation of heart rate and oxygen saturation. Invention disclosure filed with UConn. Inventors: Ki Chon and SMA Salehizadeh
3. Motion and Noise Artifact Detection and Reconstruction Algorithms for ECG Signals. Invention disclosure filed with UConn. Inventors: Ki Chon and SMA Salehizadeh
 - Invention disclosures #2 and #3 above will be licensed to a new company created by Dr. Chon and Bryant Guffey.

7. Participants and other collaborating organizations:

What individuals have worked on the project?

- Ki Chon (PI), Yitzhak Mendelson (Co-PI), Chad Darling (Co-PI) and David McManus (Co-PI)
- Jowoon Chong, Duy Dao, SMA Salehizadeh, Chae Ho Cha, Kristen Warren, Gary Zimmer, Yelena Malyuta.
- No Change

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

- Nothing to report

What other organizations were involved as partners?

- Nothing to report

8. Special Reporting Requirements

- Not applicable

9. Appendix

Seven published and accepted journal articles are provided.

Estimation of Respiratory Rates Using the Built-in Microphone of a Smartphone or Headset

Yunyoung Nam, Bersain A. Reyes, *Student Member, IEEE*, and Ki H. Chon*, *Senior Member, IEEE*

Abstract— This paper proposes accurate respiratory rate estimation using nasal breath sound recordings from a smartphone. Specifically, the proposed method detects nasal airflow using a built-in smartphone microphone or a headset microphone placed underneath the nose. In addition, we also examined if tracheal breath sounds recorded by the built-in microphone of a smartphone placed on the paralaryngeal space can also be used to estimate different respiratory rates ranging from as low as 6 breaths/min to as high as 90 breaths/min. The true breathing rates were measured using inductance plethysmography bands placed around the chest and the abdomen of the subject. Inspiration and expiration were detected by averaging the power of nasal breath sounds. We investigated the suitability of using the smartphone-acquired breath sounds for respiratory rate estimation using two different spectral analyses of the sound envelope signals: the Welch periodogram and the autoregressive spectrum. To evaluate the performance of the proposed methods, data were collected from 10 healthy subjects. For the breathing range studied (6-90 breaths/min), experimental results showed that our approach achieves an excellent performance accuracy for the nasal sound as the median errors were less than 1% for all breathing ranges. The tracheal sound, however, resulted in poor estimates of the respiratory rates using either spectral method. For both nasal and tracheal sounds, significant estimation outliers resulted for high breathing rates when subjects had nasal congestion, which often resulted in the doubling of the respiratory rates. Finally, we show that respiratory rates from the nasal sound can be accurately estimated even if a smartphone's microphone is as far as 30 cm away from the nose.

Index Terms— Respiratory rate estimation, sound intensity, tracheal sound, nasal sound, smartphone.

I. INTRODUCTION

Respiration rate (RR) is one of the key vital signs, but it is not possible to obtain it in a manner that is reliable, readily-available, cost-effective and easy to use by the general public [1]. The lack of a reliable and readily-available RR

measurement is one of the major contributors to avoidable adverse events. A retrospective study of over 14,000 cardiopulmonary arrests in acute care hospitals showed that 44% were respiratory in origin [2]. In addition, a study by Health Grades showed that respiratory failure, a key Patient Safety Indicator (PSI), has increased in U.S. Acute Care Hospitals. The reported incidence is 17.4 per 1,000 hospital admissions leading to over 15,000 avoidable deaths at a cost to the healthcare system of over \$1.8 billion [3]. Moreover, the continuous monitoring of RR as an indicator of ventilation is particularly important for patients in the intensive care unit [4].

The most common method for measuring RR consists of either manually counting the chest wall movements or auscultation of breath sounds with a stethoscope. Previous studies have shown that these manual methods tend to be unreliable in acute care settings and are limited by their intermittent cadence [5]. For automated approaches to RR assessment, sensors that measure airflow are often used in clinical settings. The airflow is usually measured by spirometry devices and some of the popular sensors include pneumotachograph or nasal cannulae that are connected to a pressure transducer, heated thermistor, or anemometer. Although the spirometry devices provide accurate estimates of RR, breathing through a mouthpiece or facemask connected to a pneumotachograph is inconvenient and adds unnecessary airway resistance. More importantly, due to high cost, inconvenience for patients' everyday use, and immobility of the traditional spirometry devices, there is impetus for developing simple, cost-effective and portable devices for estimating RR [4].

One approach has the potential to meet the above criteria for easily-accessible, affordable and on-demand monitoring of RR via the use of smartphones without any external sensors. We have recently shown that good estimates of resting RR can be obtained directly from a finger's pulsatile light intensity fluctuations which are captured using the smartphone's built-in camera [6]. However, the accuracy of this approach for RR estimation degrades when breathing rates are higher than 30 breaths/min. To mitigate this limitation, we propose in this work to use either a built-in microphone or the microphone of a headset plugged to a smartphone to estimate RR over a wide dynamic range.

The stethoscope is a common device routinely used by physicians to determine the health status of the respiratory system. Given that a stethoscope is essentially a microphone, it should be expected that a microphone can also be used to obtain

This work was supported by the Soonchunhyang University Research Fund and was also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2014R1A1A1004033).

Asterisk indicates corresponding author.

Yunyoung Nam is with the Department of Computer Science and Engineering, Soonchunhyang University, Asan 336-745, Korea (e-mail: ynam@sch.ac.kr).

B. A. Reyes and K. H. Chon are with the Department of Biomedical Engineering, University of Connecticut, Storrs, CT, 06269 USA (e-mail: {bersain.reyes, kchon}@uconn.edu).

RR [7], [8]. A key technical challenge is to discriminate between inspiratory and expiratory sound signals so that a correct RR can be determined. Fortunately, dynamics of the inspiration and expiration are different, hence, many different approaches can be used [9]–[12] to discriminate between the two phases of the breathing cycle. Some of the noted automated approaches for respiratory rate estimation include estimating the intensity changes of breathing sounds [13], the relative changes of the total sound power [14], the analysis of the tracheal sound entropy [15], and bioacoustic analysis [10]. While expiratory sounds recorded at the trachea are slightly louder than inspiratory sounds [16], in general the dynamic characteristics of inspiration are similar to expiration sounds recorded at the trachea. In contrast, the intensities of nasal breath sounds recorded near a subject's nose during inspiration and expiration are different. Hence, by taking advantage of different acoustic properties of respiration sounds measured at either trachea or nose, and with a built-in smartphone microphone or the microphone of a headset cabled to a smartphone, we developed an approach to estimate RR that is accurate for a wide dynamic range from 6 breaths/min to 90 breaths/min. The aim of the paper is to show that our method allows a reliable determination of the RR without any external sensors but by utilizing only the built-in microphone or headset microphone of a smartphone.

II. MATERIALS AND METHODS

A. Data acquisition

Data were collected from ten healthy non-smoking volunteers with ages ranging from 20 to 40 years. The experimental protocol was approved by the Institutional Review Board of Worcester Polytechnic Institute and all volunteers signed the informed consent prior to data acquisition.

Data were collected while volunteers were seated in the upright position in a regular office room. Tracheal and nasal breathing sound signals were recorded by a built-in microphone and the headset microphone of an iPhone 4S (Apple, Inc., Cupertino, CA, USA) placed on the subject's suprasternal notch on the neck and the philtrum under the nose, respectively. For tracheal breath sound recording, the built-in microphone was manually kept in a fixed position by the subject, while for nasal breathing sound recording, the headset's microphone was placed gently under the subject's nose to assure that it would not be displaced during the experiment. None of the subjects reported any discomfort with the microphone placement.

For determining true breathing rates, inductance plethysmography bands were placed around the subject's chest and abdomen (Respirace, Ambulatory Monitoring, Inc., Ardsley, NY, USA). These reference signals were acquired and stored in a personal computer using LabChart 7 software (ADInstruments, Inc., Dunedin, New Zealand) at a sampling rate of 400 Hz. Breathing sound data from microphones were collected directly to an iPhone using 16-bits per sample with a sampling rate of 44,100 Hz. In addition to collecting sound signals, the amplitude of their envelope was also computed and

stored on the smartphone. First, the audio signals were bandpass filtered in the range of 500-5000 Hz to remove the effects of low-frequency and high-frequency noise. The Hilbert transform was used to extract the envelope of the filtered sound signal. For a continuous-time signal $x(t)$, its Hilbert transform is defined as follows [17]:

$$H(x(t)) = \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \frac{1}{t-\tau} d\tau. \quad (1)$$

The amplitude of the envelope was calculated as the magnitude of its analytic signal, i.e. complex valued. The resulting signal was then digitized and stored in the smartphone for further processing offline. The amplitude of the envelope was digitized at a rate of 100 Hz. These envelope signals were digitized at this lower rate to reduce computational time and data capacity, and mainly because the highest breathing rate we were concerned with was at most 2 Hz.

All subjects were instructed to breathe at a metronome rate according to a timed beeping sound programmed at a given frequency. Each subject was wearing earphones and was asked to inhale at each beep sound followed by exhalation before the next beep sound occurred. Data were collected for breathing frequencies ranging from 0.1 to 1.5 Hz at increments of 0.1 Hz, which corresponds to breathing rates ranging from 6 to 90 breaths/min at steps of 6 breaths/min. Prior to data collection, all subjects were acclimated to the different metronome breathing rates. For each subject, three minutes of breathing data were collected at each programmed metronome frequency. For breathing rates greater than 60 breaths/min, subjects were given ample time break before the start of the next breathing rates.

B. Preprocessing

The smartphone's sampling frequency was not constant but varied around 100 Hz, therefore, a cubic spline algorithm was used to interpolate the digitized signals to a constant 100 Hz. Finally, the smoothed envelope signals were band-pass filtered between 0.19 and 4.6 Hz, using a rectangular window in the frequency domain, and then down-sampled from 100 Hz to 10 Hz. After this down-sampling, the initial and final 10 seconds of the recordings were discarded. This preprocessing was performed on a personal computer using Matlab (R2012a, The Mathworks, Inc., Natick, MA, USA).

C. Data Analysis

In order to determine the appropriate respiratory rate, the power spectrum of the pre-processed tracheal and nasal sound envelopes were investigated using two different methods: 1) the Welch periodogram technique, and 2) the autoregressive (AR) power spectral analysis technique via the Burg algorithm. In the first approach, the PSD of each segment was computed via Welch periodogram. In the second approach, the Burg method was used to estimate the AR coefficients from sampled data by the simultaneous minimization of the forward and backward linear prediction squared errors, while the AR coefficients were constrained to satisfy the Levinson-Durbin recursion. An AR model order of 50 was employed based on

the minimum description length criterion [18].

The respiratory rate can be calculated by either the largest peak of PSDs of the sound envelope or via the AR model of a sound envelope. The true respiration rate was found by computing the PSD of the reference respiration signal and finding the frequency at which its maximum spectral peak occurred. In general, the average error of the respiratory rate estimation based on the maximum spectral peak of PSD was larger for the high frequency breathing rates. The respiratory frequency can be determined by the frequency corresponding to the maximum peak of PSD, provided that the frequency spectra of inspiration and expiration are similar. However, they become different when subjects have nasal congestion, for example. When this occurs, the respiratory rates can be (incorrectly) double.

The estimated RR from the smartphone's microphone was compared to the true RR acquired from the Respirace system to test the reliability and accuracy of the proposed method. For each respiratory rate, detection errors were found for all subjects using the two different spectral methods as previously described. The respiratory rate estimation error ε was calculated for each respiratory frequency:

$$\varepsilon = \frac{m \text{ean}(R - R_{\text{est}})^2}{\text{mean}(R)^2} \times 100, \quad (2)$$

where R and R_{est} represent reference and estimated respiratory rate, respectively. The values of error were averaged for all subjects.

Fig. 1 shows the block diagram of the proposed method for estimating the respiratory rate from either the tracheal or nasal breath sounds acquired from a smartphone.

III. RESULTS

Fig. 2 shows a typical 20-second recording with a built-in microphone and an earpiece (headset) microphone together with their corresponding sound spectrograms. Fig. 2 (a) and (b) show the raw data of the tracheal and nasal breath sounds, respectively. Fig. 2 (c) shows the sound spectrogram of tracheal and nasal breathing signals, where the vertical axis corresponds to the frequency and the horizontal axis to the time. Each color represents the power of the signal at a specific time and frequency, with the power decreasing from red to blue. For the tracheal sounds, inspiration and expiration tend to have similar frequency distributions. Note that more spectral power

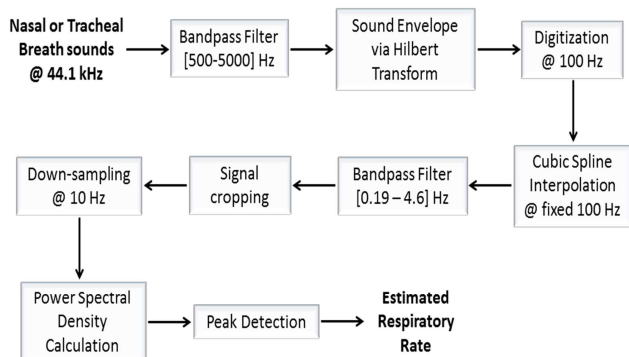


Fig. 1. Flowchart of the proposed method for respiratory rate estimation using smartphone acquired tracheal and nasal breath sounds.

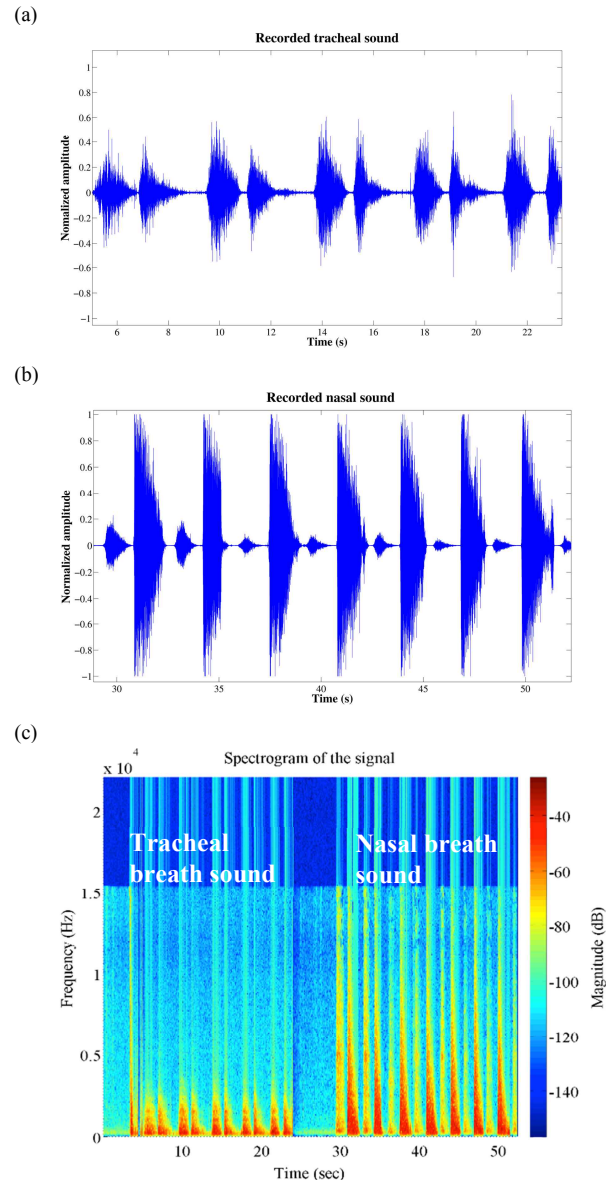


Fig. 2. Examples of recorded tracheal and nasal breath sound. (a) Recorded tracheal signal using the built-in microphone. (b) Recorded nasal signal using the earpiece microphone. (c) Sound spectrogram.

is observed in nasal breath sound than in tracheal breath sound. Also, the spectral power of inspiratory phases in nasal breath sound is less than that of their expiratory phases.

Examples of estimated envelope signals from recorded tracheal and nasal sounds are presented in Fig. 3. Typical unfiltered envelope signals from a built-in smartphone microphone and an earpiece-type headset's microphone are shown respectively in Fig. 3 (a) and (c) while the post-bandpass filtered and cubic spline interpolated envelope signals are shown in Fig. 3 (b) and (d). Note that undesired sound activities are removed from the original signals. It can be seen that the filtered and splined magnitude signals follow the absolute values of flow signals, as has been reported by other authors [19]. Note that the amplitude of the estimated flow does not represent the actual amount of flow in liters per second, as it is not calibrated.

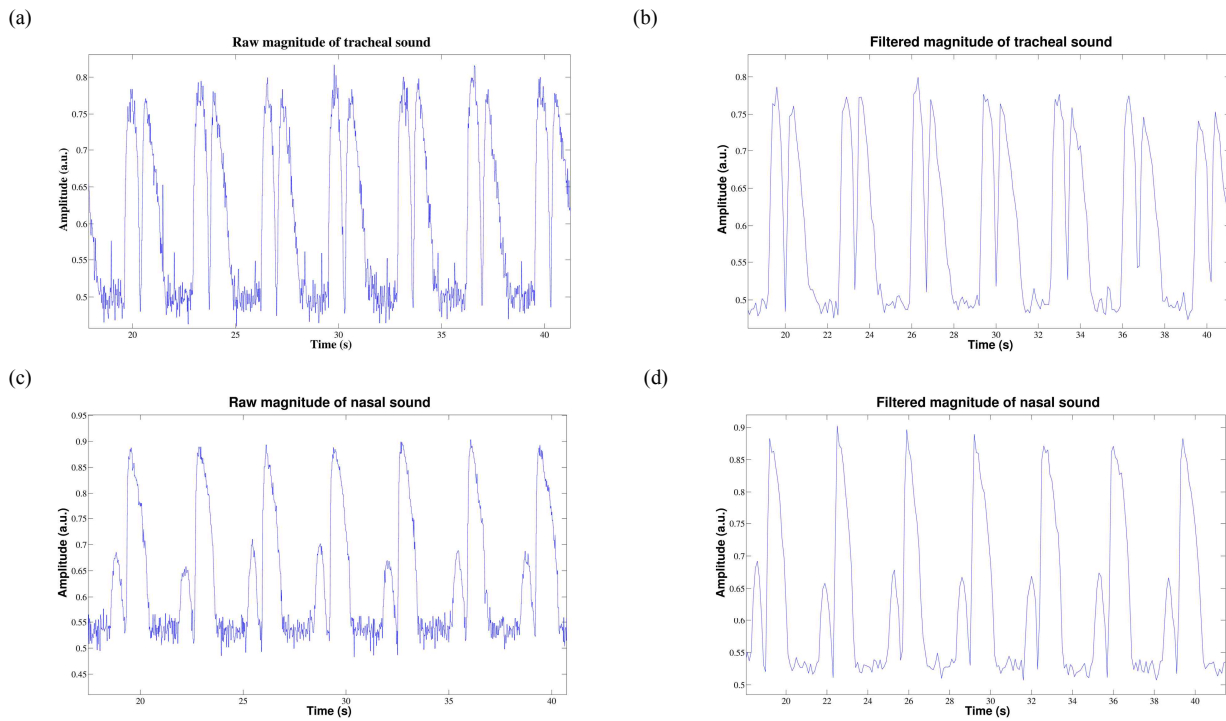


Fig. 3. Typical recorded tracheal and nasal sound envelope signals. (a) Raw envelope of tracheal sound. (b) Filtered envelope of tracheal sound signal. (c) Raw envelope of nasal sound. (d) Filtered envelope of nasal sound.

Fig. 4 shows two different spectral estimates of the filtered sound envelope signals: the standard power spectrum and the AR model-based spectrum from a normal subject, shown in Fig. 4 (a), and from a subject suffering nasal congestion, shown in Fig. 4 (b), both when the breathing rate was 0.3 Hz. In these figures, the maximum peaks were obtained at the first and the second order harmonic frequencies, respectively. In general, the respiratory frequency can be determined as the frequency corresponding to the maximum peak of the PSD. However, the breathing rate of a subject suffering from nasal congestion was found to be doubled, as shown in Fig. 4 (b).

Fig. 5 shows boxplots with the median and interquartile range (IQR) errors obtained from the respiratory rates calculated by the maximum peak in PSDs of tracheal sound envelope and AR spectrum of the tracheal sound envelope, shown in Fig. 5 (a)-(b), and the PSD of the nasal sound envelope and the AR spectrum of the nasal sound envelope, shown in Fig. 5 (c)-(d). These figures indicate how well these two methods perform in estimating respiratory rates. The median and IQR of respiratory rate estimation errors were obtained from each reference and derived respiration rate as defined in Eq. (2). The lower boundary of the box indicates the 25th percentile, a line within the box marks the median, and the upper boundary of the box 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 sample. Red crosses represent the outliers.

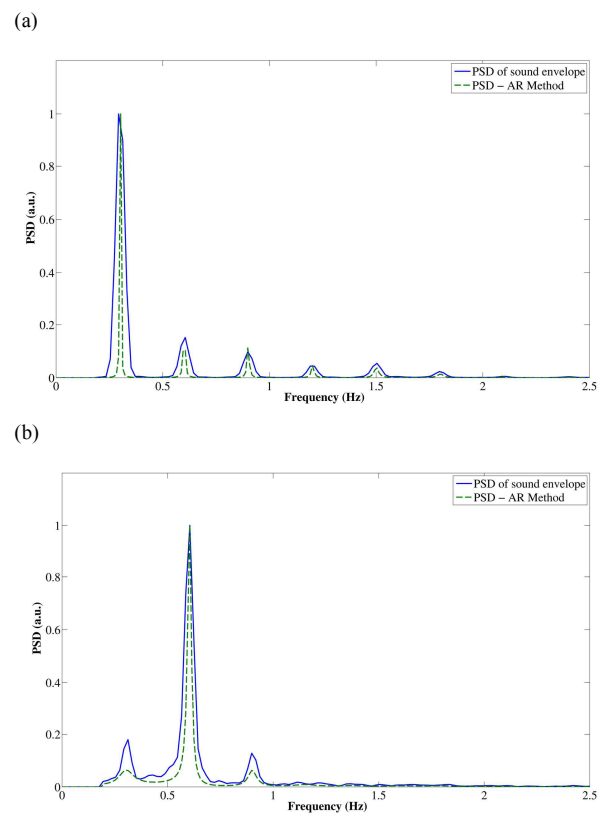


Fig. 4. PSD examples of nasal sounds using envelope and AR model approaches. (a) Normal nasal breathing at frequency rate of 0.3 Hz. (b) Nasal breathing with nasal obstruction at frequency rate of 0.3 Hz.

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

5

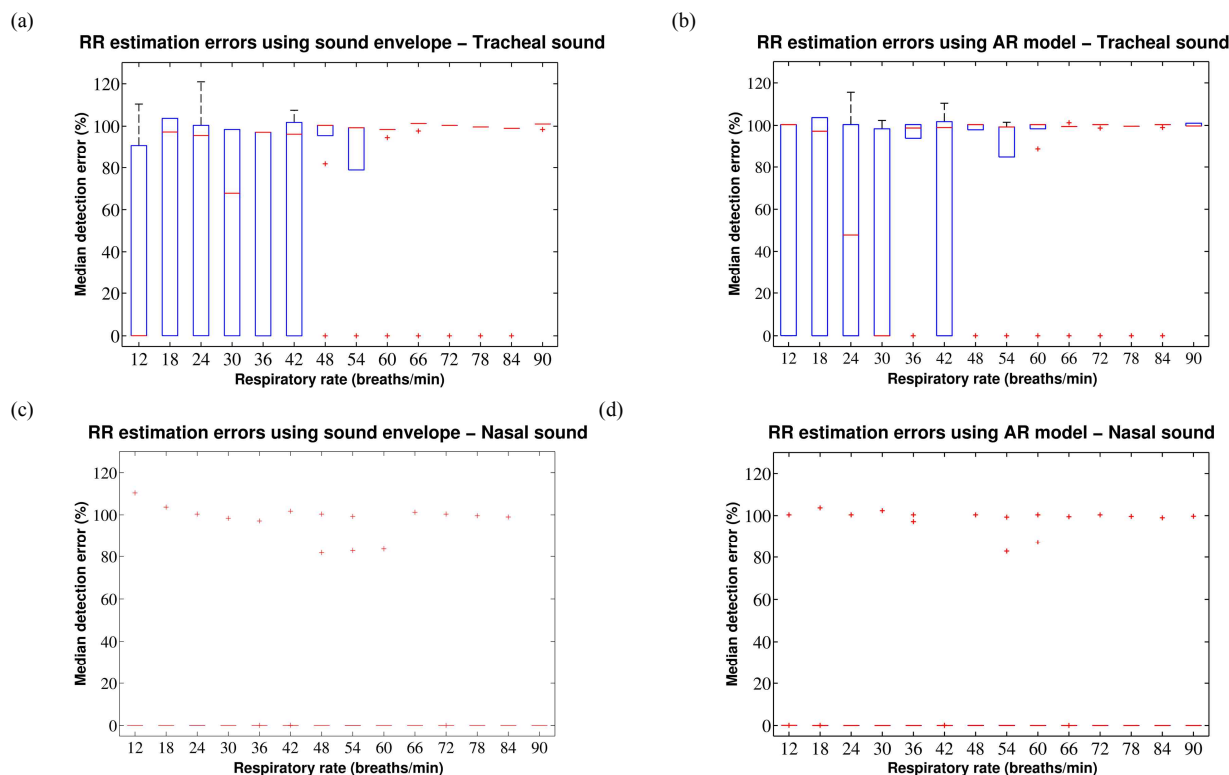


Fig.5. Median and IQR respiratory rate estimation errors ϵ measured from the respiratory rate results calculated by the maximum peak in PSDs from (a) tracheal breath sound, (b) AR model of tracheal breath sound, (c) nasal breath sound, and (d) AR model of nasal breath sound.

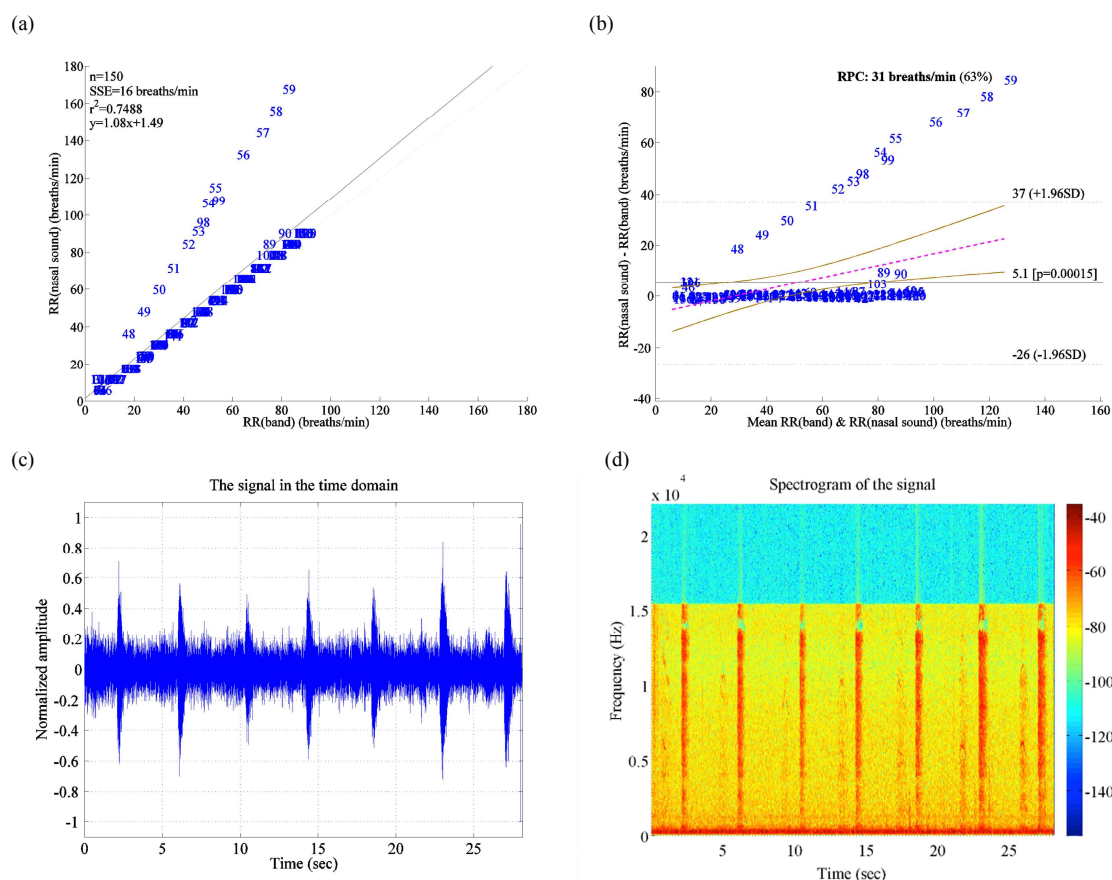


Fig. 6. Examples of Bland-Altman plot, correlation plot, and the recorded nasal breathing sound recorded with a built-in microphone of an iPhone 4S. (a) An example of a correlation plot. (b) An example of a Bland-Altman plot with proportional bias regression line. (c) Time waveform. (d) Spectrogram.

The respiratory rate estimation error ϵ was found to be lowest for nasal sound envelope and AR model from nasal sound envelope at all breathing rates as shown in Table I. The errors ϵ of sound envelope and AR model of nasal sound envelope were 7.43% and 0.025%, respectively. There was no significant difference in the average respiratory rate estimation error between the two approaches except for 0.1 Hz. As shown in

Table I (RR=0.1 Hz), it is difficult to detect low frequency component (envelope) because the signal-to-noise ratio decreases sharply at the low frequency band.

Fig. 6 shows the correlation and Bland-Altman plots with experiment numbers (1-150) for the mean RR data from nasal breathing signals and the inductance plethysmography bands, a typical 30-second nasal breathing sound signal obtained with

Table I. Accuracy as determined by median errors and IQR measured from the respiratory rate results obtained from nasal breathing sound signal (N=10).

Respiration Rate Breaths/min (beep sound)	Respiration Rate Breaths/min (bands)	Estimation Error ϵ	Sound envelope		AR model of sound envelope	
			Tracheal sound	Nasal sound	Tracheal sound	Nasal sound
6 (0.1 Hz)	6.055(0.101 Hz)	Median	90.821±18.775	111.306±95.292	58.518±136.685	0.3±42.535
		IQR	40.969±19.313	119.823±52.329	85.394±35.133	90.768±42.73
12 (0.2 Hz)	13.184 (0.22 Hz)	Median	97.148±45.796	0.055±33.108	29.909±13.734	0.064±40.132
		IQR	96.241±46.047	0±0	98.998±46.61	25.084±11.824
18 (0.3 Hz)	18.164 (0.303 Hz)	Median	101.792±46.731	0.055±31.087	101.792±47.34	0.008±31.1
		IQR	100.391±29.857	0±0	100.391±39.924	0.012±0.005
24 (0.4 Hz)	23.965 (0.399 Hz)	Median	8.278±3.903	0.055±30.1	26.921±12.691	0±30.117
		IQR	99.307±38.951	0.01±0.004	99.307±39.146	0±0
30 (0.5 Hz)	29.883 (0.498 Hz)	Median	40.163±18.933	0.024±29.526	35.748±16.852	0.002±30.707
		IQR	98.444±29.427	0±0	100.391±29.794	0±0
36 (0.6 Hz)	35.918 (0.599 Hz)	Median	0.959±0.452	0.008±29.143	4.31±2.032	0.005±39.514
		IQR	101.282±30.284	0.012±0.005	99.503±29.915	24.29±11.45
42 (0.7 Hz)	42.305 (0.705 Hz)	Median	0.886±0.417	0.002±30.534	0±0	0.002±0.031
		IQR	100.391±30.116	0.026±0.012	100.391±30.067	0.002±0.001
48 (0.8 Hz)	47.93 (0.799 Hz)	Median	0±0	0±36.67	0.406±0.191	0±40.155
		IQR	99.64±29.891	20.444±9.637	99.64±29.892	25.107±11.836
54 (0.9 Hz)	53.73 (0.896 Hz)	Median	0±0	0±36.598	0±0	0±36.6
		IQR	98.998±29.699	20.731±9.773	100.391±30.074	20.701±9.759
60 (1.0 Hz)	59.766 (0.996 Hz)	Median	0±0	0.002±25.065	0.348±0.164	0.002±37.626
		IQR	101.044±1.04	0±0	99.74±0.598	21.791±10.272
66 (1.1 Hz)	66.152 (1.103 Hz)	Median	0.65±0.307	0.003±30.384	1.305±0.615	0.001±29.851
		IQR	97.148±45.796	0±0	29.909±13.734	0.001±0
72 (1.2 Hz)	72.072 (1.201 Hz)	Median	96.241±46.047	0.005±30.116	98.998±46.61	0±30.117
		IQR	101.792±46.731	0.001±0	101.792±47.34	0±0
78 (1.3 Hz)	77.93 (1.299 Hz)	Median	100.391±29.857	0.004±29.891	100.391±39.924	0±29.892
		IQR	8.278±3.903	0±0	26.921±12.691	0±0
84 (1.4 Hz)	83.79(1.397 Hz)	Median	99.307±38.951	0.002±29.699	99.307±39.146	0.001±29.699
		IQR	40.163±18.933	0±0	35.748±16.852	0±0
90 (1.5 Hz)	90.234 (1.504 Hz)	Median	98.444±29.427	0.001±0	100.391±29.794	0.001±39.896
		IQR	0.959±0.452	0±0	4.31±2.032	24.935±11.755

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

7

the built-in microphone of an iPhone 4S, and its corresponding sound spectrogram. Similarly, Fig. 7 shows the raw data of the recorded nasal breathing sound during spontaneous breathing when the distance between the nose and iPhone was 30 cm.

In Fig. 7 (a) and (b), inspiration and expiration were observed in both the sound signal and the magnitude signal. Fig. 7 (c) contains peaks near 0.2539 Hz, 0.4883 Hz, and 0.7227 Hz. The real breathing rate was 0.2539 Hz. Finally, Fig. 8 shows examples of the recorded nasal breathing sound with background vocal noises during spontaneous breathing when the distance between nose and iPhone was 30 cm. Respiratory rate was measured by a peak near 0.3125 Hz as shown in Fig. 8 (c). The real breathing rate was 0.3125 Hz as measured by the Resptrace system.

IV. DISCUSSION AND CONCLUSION

In this paper, methods for estimating respiratory rate from tracheal and nasal breathing sound signals have been presented.

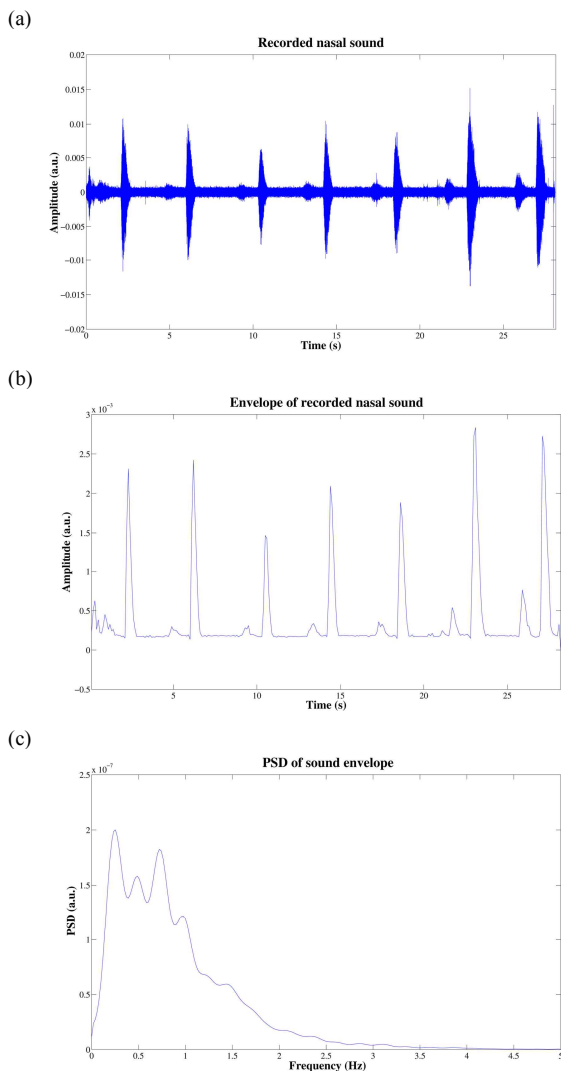


Fig. 7. Example of recorded nasal breath sound when the distance between the subject's nose and iPhone was 30 cm. (a) Filtered sound signal. (b) Down-sampled envelope signal. (c) PSD of the envelope.

Previously, our research group explored the feasibility of using a smartphone together with a specifically designed acoustical sensor to record tracheal sounds from which respiratory rates could be estimated [20]. In contrast, the feasibility of using smartphones together with their built-in and standard headset microphones for estimating respiratory rates was tested in this paper. Another motivation for this work is based on several previous studies that showed that accurate respiratory rates, especially at low breathing rates, could be obtained from pulse oximeters, but the accuracy of this approach degrades above 30 breaths/min. Theoretically, the characteristics of the breathing sounds obtained from smartphones' microphones match the respiratory rate; thus, accurate respiratory rates can be obtained. Our results indicate that certainly for low and high breathing ranges (0.1 - 1.5 Hz), this is feasible from nasal breathing sounds recorded from smartphone microphones. For 0.1 Hz,

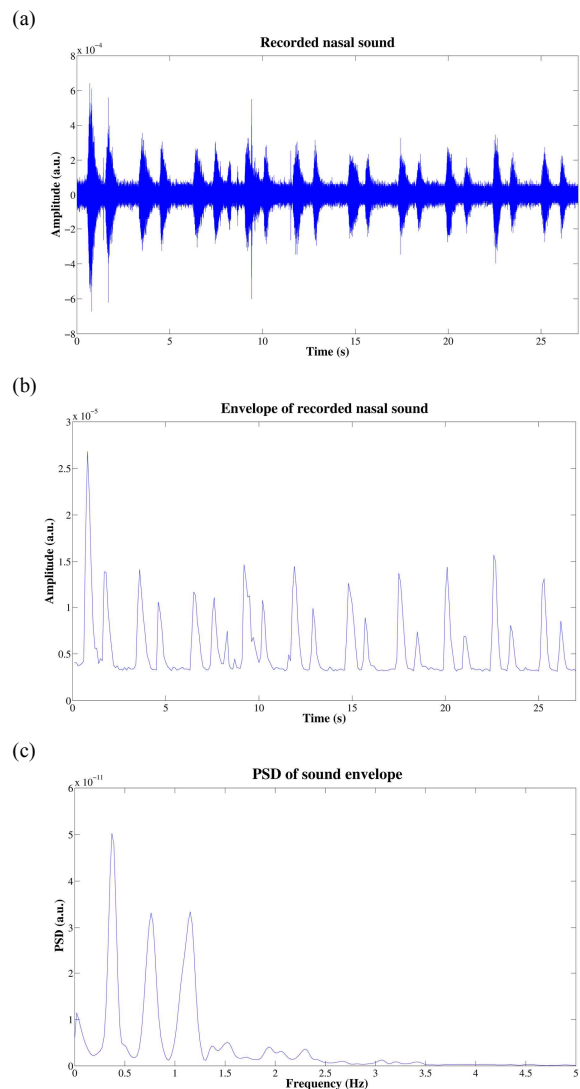


Fig. 8. Example of recorded nasal breath sound with background vocal noises when the distance between the subject's nose and iPhone was 30 cm. (a) Filtered sound signal. (b) Down-sampled envelope signal. (c) PSD of the envelope.

due to randomly occurring background noise or other noise during acquisition, RR estimation could not be reliably estimated. For reliable estimates of RR, the background noise should be kept at minimal which can be problematic for intensive care units.

We compared both Welch and AR spectra of the tracheal and nasal envelope sounds acquired with smartphones for respiratory rate estimation using the largest spectral peaks. Both spectral methods provided accurate respiratory rate estimation from nasal sounds in this study for both low and high breathing rates. However, for high breathing rates (0.8 - 1.5 Hz), a simple approach using the largest spectral peak detection could not always provide good results especially when subjects suffered from nasal congestion.

Microphone sensitivity is typically measured with a 1 kHz sine wave at a 94 decibels (dB) sound pressure level (SPL), or 1 pascal (Pa) pressure. The magnitude of the analog or digital output signal from the microphone with that input stimulus is a measure of its sensitivity. In this work, the sound signals were obtained by an iPhone 4S that has 2 microphones, a Infineon 1014 microphone on top of the device and a Knowles S1950 microphone in the bottom [21]. The Infineon 1014 microphone is used for canceling out background noise, and it is located on the top of the unit near the headphone jack; the main microphone is on the bottom left [22]. All current iOS devices (iPhone 3GS and later, iPod touch 4 and later, and all iPads) include built-in microphones. However, Apple included a very steep high-pass filter, which presumably works as a wind and pop filter. The low-frequency roll-off for the internal microphone in these devices is very steep, on the order of 24 dB / octave starting at 250 Hz [23]. However, with the advent of the iOS 6, we are able to turn off the low-frequency roll off filter, thereby resulting in fairly flat response [23]. Even though the performance of a smartphone is limited, we have compensated for these microphones as much as possible.

Better performance in detecting the apnea-hypopnea index (AHI) or sleep apnea/hypopnea syndrome (SAHS) from analysis of breath sounds recorded by a microphone of a smartphone can be achieved when combined with an oximetry signal. There have been efforts to monitor patients with asthma and chronic obstructive pulmonary disease [24], [25], with other severe respiratory diseases [26], and with nasal obstruction [27], however there have not been studies reporting respiratory rate estimation considering nasal congestion, to our knowledge. In this paper, the spectral morphology of nasal sound signals was analyzed to develop the respiratory rate estimation methods. The intensity changes of nasal sound signals were investigated to choose the best approach.

Since some people may feel some discomfort when they need to use an earpiece-microphone or placing a smartphone's microphone directly underneath their nose, a non-contact breathing sound acquisition has been conducted to illustrate that breathing rates can still be accurately derived even if a smartphone is 30 cm away from the nose. Certainly, this approach is more prone to background noise, however, provided that they are not overwhelming and mask the nasal breathing sounds, our approach appears to provide good

respiratory rate estimates. A more thorough analysis under various background noise levels will need to be performed to augment our preliminary results on having a smartphone as far as 30 cm from the nose.

In summary, we have shown that accurate RR can be estimated for nasal breathing sounds acquired from a smartphone's microphone as the estimation errors were less than 1% for cases considered in this work. The fact that our approach does not require any external sensors, as all that is required is to place a smartphone's microphone underneath one's nose, is attractive from many perspectives. To date, with promising results obtained from this work, smartphones can be used as a vital sign monitoring device that can readily provide heart rates and respiratory rates rather reliably without using any expensive external sensors. It is expected that future work by either our laboratory or others will result in additional vital sign capabilities directly from smartphones' or tablets' resident sensors.

REFERENCES

- [1] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign," *Med. J. Aust.*, vol. 188, no. 11, p. 657, 2008.
- [2] M. A. Peberdy, W. Kaye, J. P. Ornato, G. L. Larkin, V. Nadkarni, M. E. Mancini, R. A. Berg, G. Nichol, and T. Lane-Trullt, "Cardiopulmonary resuscitation of adults in the hospital: a report of 14 720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation," *Resuscitation*, vol. 58, no. 3, pp. 297-308, 2003.
- [3] HealthGrades, "The Sixth Annual HealthGrades Patient Safety in American Hospitals." 2009.
- [5] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatr. Pulmonol.*, vol. 46, no. 6, pp. 523-529, Jun. 2011.
- [5] K. Hillman, J. Chen, M. Cretikos, R. Bellomo, D. Brown, G. Doig, S. Finfer, A. Flabouris, and others, "Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial," *Lancet*, vol. 365, no. 9477, pp. 2091-2097, 2005.
- [6] Y. Nam, J. Lee, and K. Chon, "Respiratory Rate Estimation from the Built-in Cameras of Smartphones and Tablets," *Ann. Biomed. Eng.*, vol. 42, no. 4, pp. 885-898, 2014.
- [7] G. Comtois, J. I. Salisbury, and Y. Sun, "A smartphone-based platform for analyzing physiological audio signals," in *Bioengineering Conference (NEBEC), 2012 38th Annual Northeast*, 2012, pp. 69-70.
- [8] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds: advances beyond the stethoscope," *Am. J. Respir. Crit. Care Med.*, vol. 156, no. 3, pp. 974-987, 1997.
- [9] J. S. Chuah and Z. K. Moussavi, "Automated respiratory phase detection by acoustical means," *Univ. Mani. Dept Electr. Comput. Eng.*, 2004.
- [10] P. Hult, B. Wranne, and P. Ask, "A bioacoustic method for timing of the different phases of the breathing cycle and

- monitoring of breathing frequency,” *Med. Eng. Phys.*, vol. 22, no. 6, pp. 425–433, 2000.
- [11] A. Kulkas, E. Huupponen, J. Virkkala, M. Tenhunen, A. Saastamoinen, E. Rauhala, and S.-L. Himanen, “Intelligent methods for identifying respiratory cycle phases from tracheal sound signal during sleep,” *Comput. Biol. Med.*, vol. 39, no. 11, pp. 1000–1005, 2009.
- [12] Z. K. Moussavi, M. T. Leopando, H. Pasterkamp, and G. Rempel, “Computerised acoustical respiratory phase detection without airflow measurement,” *Med. Biol. Eng. Comput.*, vol. 38, no. 2, pp. 198–203, 2000.
- [13] J. Cummiskey, T. Williams, P. Krumpe, and C. Guilleminault, “The detection and quantification of sleep apnea by tracheal sound recordings,” *Am. Rev. Respir. Dis.*, vol. 126, no. 2, pp. 221–224, 1982.
- [14] H. Nakano, M. Hayashi, E. Ohshima, N. Nishikata, T. Shinohara, and others, “Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome,” *SLEEP-N. Y. THEN Westchest.*, vol. 27, no. 5, pp. 951–958, 2004.
- [15] A. Yadollahi and Z. M. Moussavi, “A robust method for estimating respiratory flow using tracheal sounds entropy,” *Biomed. Eng. IEEE Trans. On*, vol. 53, no. 4, pp. 662–668, 2006.
- [16] M. Mahagnah and N. Gavriely, “Repeatability of measurements of normal lung sounds,” *Am. J. Respir. Crit. Care Med.*, vol. 149, no. 2, pp. 477–481, 1994.
- [17] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.
- [18] S. Haykin, *Adaptive Filter Theory* (2Nd Ed.). Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1991.
- [19] C.-L. Que, C. Kolmaga, L.-G. Durand, S. M. Kelly, and P. T. Macklem, “Phonspirometry for noninvasive measurement of ventilation: methodology and preliminary results,” *J. Appl. Physiol. Bethesda Md* 1985, vol. 93, no. 4, pp. 1515–1526, Oct. 2002.
- [20] B. A. Reyes, N. Reljin, and K. H. Chon, “Tracheal Sounds Acquisition Using Smartphones,” *Sensors*, vol. 14, no. 8, pp. 13830–13850, Jul. 2014.
- [21] “iPhone 4 Microphone Teardown,” <http://www.ifixit.com/Teardown/iPhone+4+Microphone+Teardown/3473>.
- [22] Wikipedia, “iPhone 4S.” Wikipedia, the free encyclopedia, 2011.
- [23] Faber Acoustical, “iOS 6 kills the filter on headset and mic inputs,” <http://blog.faberacoustical.com/2012/ios/iphone/finally-ios-6-kills-the-filter-on-headset-and-mic-inputs>, 2013.
- [24] M. Hasselgren, 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*, vol. 19, no. 1, pp. 54–57, 2001.
- [25] D. C. Willems, M. A. Joore, J. J. Hendriks, R. A. van Duurling, E. F. Wouters, and J. L. Severens, “Process evaluation of a nurse-led telemonitoring programme for patients with asthma,” *J. Telemed. Telecare*, vol. 13, no. 6, pp. 310–317, 2007.
- [26] C. Maiolo, E. I. Mohamed, C. M. Fiorani, and A. De Lorenzo, “Home telemonitoring for patients with severe respiratory illness: the Italian experience,” *J. Telemed. Telecare*, vol. 9, no. 2, pp. 67–71, 2003.
- [27] H. Choi, I.-H. Park, H. G. Yoon, and H.-M. Lee, “Wireless patient monitoring system for patients with nasal obstruction,” *Telemed. E-Health*, vol. 17, no. 1, pp. 46–49, 2011.

Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Physiol. Meas. 36 2317

(<http://iopscience.iop.org/0967-3334/36/11/2317>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 137.99.13.188

This content was downloaded on 09/10/2015 at 16:33

Please note that [terms and conditions apply](#).

Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals

Jesús Lázaro^{1,2}, Yunyoung Nam³, Eduardo Gil^{1,2},
Pablo Laguna^{1,2} and Ki H Chon⁴

¹ BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, Universidad de Zaragoza, 12, 50018, Zaragoza, Spain

² Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain

³ Department of Computer Science and Engineering, Soonchunhyang University, Asan 336–745, Korea

⁴ Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269-3247, USA

E-mail: jlazarop@unizar.es, yynams@gmail.com and kchon@engr.uconn.edu

Received 24 April 2015, revised 16 August 2015

Accepted for publication 21 August 2015

Published 9 October 2015



CrossMark

Abstract

A method for deriving respiratory rate from smartphone-camera-acquired pulse photoplethysmographic (SCPPG) signal is presented. Our method exploits respiratory information by examining the pulse wave velocity and dispersion from the SCPPG waveform and we term these indices as the pulse width variability (PWV). A method to combine information from several derived respiration signals is also presented and it is used to combine PWV information with other methods such as pulse amplitude variability (PAV), pulse rate variability (PRV), and respiration-induced amplitude and frequency modulations (AM and FM) in SCPPG signals.

Evaluation is performed on a database containing SCPPG signals recorded from 30 subjects during controlled respiration experiments at rates from 0.2 to 0.6 Hz with an increment of 0.1 Hz, using three different devices: iPhone 4S, iPod 5, and HTC One M8. Results suggest that spontaneous respiratory rates (0.2–0.4 Hz) can be estimated from SCPPG signals by the PWV- and PRV-based methods with low relative error (median of order 0.5% and interquartile range of order 2.5%). The accuracy can be improved by combining PWV and PRV with other methods such as PAV, AM and/or FM methods. Combination of these methods yielded low relative error for normal respiratory rates, and

maintained good performance at higher rates (0.5–0.6 Hz) when using the iPhone 4S or iPod 5 devices.

Keywords: respiration, photoplethysmography, PPG, pulse width variability, PWV

(Some figures may appear in colour only in the online journal)

1. Introduction

Monitoring of respiration is usually performed by techniques such as spirometry, pneumography, and plethysmography. These techniques require cumbersome devices which are impractical in certain situations such as stress test or sleep studies (Bailón *et al* 2006b), and which may interfere with natural breathing. Thus, obtaining accurate respiratory information from comfortable non-invasive devices is a task of interest.

This paper is focused on deriving respiratory rate by using smartphone devices. This cannot fully replace spirometry which offers also information about respiratory-volume-related parameters. However, respiratory rate by itself is useful in several situations, e.g. it remains a sensitive clinical parameter in many pulmonary diseases (Krieger *et al* 1986) such as acute respiratory dysfunction (Gravelyn and Weg 1980). Impedance-pneumography-based techniques can be used when respiratory rate is the only respiratory information required since it is not designed to obtain other physiological parameters. These techniques are non-invasive and comfortable as they use only a pair of electrodes to measure the impedance changes in the chest. However, they often lead to unusable signals due to low signal-to-noise ratio and motion artifacts (Larsen *et al* 1984).

Many algorithms for deriving respiratory rate from comfortable non-invasive devices have been presented. Most of them use the electrocardiogram (ECG), exploiting variations of beat morphology and/or occurrence (Mason and Tarassenko 2001, Bailón *et al* 2006a, 2006b, Lázaro *et al* 2014a). There are methods based on other biomedical signals such as blood pressure (De Meersman *et al* 1996), photoplethysmographic (PPG) signals (Chon *et al* 2009, Lázaro *et al* 2013), and pulse transit time (Chua and Heneghan 2005), which require both ECG and PPG signals to derive respiratory rate.

The PPG signal is usually provided by a biomedical sensor called a pulse oximeter. It is composed of a light source which illuminates tissue (usually fingers, earlobes, or forehead) and a light detector which measures the reflected or transmitted light depending on its position, leading to a signal which is proportional to the blood volume. Deriving respiratory information from a PPG signal is particularly interesting, because pulse oximeters are very simple, economical, and comfortable to use. Furthermore, the pulse oximeter is widely adopted to monitor the peripheral oxygen saturation, which constitutes a very relevant parameter in the study of respiration. Thus, the pulse oximeter is a very valuable device in clinical settings.

Known methods for deriving respiratory rate from the PPG signal exploit variations on pulse morphology and/or occurrence. It is well known that respiration modulates heart rate (Hirsch and Bishop 1981) leading to a respiratory component in heart rate variability (Task Force 1996), which is also seen in pulse rate variability (PRV) since they are highly correlated (Gil *et al* 2010). Respiration also modulates the morphology of the PPG signal. Inspiration can lead to a reduction in tissue blood volume, and this lowers the amplitude of the PPG signal. This reduction in tissue blood volume is generated by two different mechanisms:

a reduction of cardiac output, and a reduction of intra-thoracic pressure (Meredith *et al* 2012). Variations in amplitude of the PPG signal have been used to obtain respiratory information (Johansson and Oberg 1999), and both heart and respiratory rates were extracted by methods based on empirical mode decomposition (Garde *et al* 2013), and based on correntropy spectral density (Garde *et al* 2014). There have been proposed other methods for obtaining respiratory rate based on the respiration-related amplitude and frequency modulations (AM and FM, respectively) in PPG signal (Chon *et al* 2009). Also, pulse width variability (PWV) have been proposed for deriving respiratory rate, either alone or in combination with other methods such as pulse amplitude variability (PAV) and PRV (Lázaro *et al* 2013). A time-frequency-coherence-based combination of PWV, PAV and PRV have been also proposed (Peláez-Coca *et al* 2013).

Smartphone devices can record PPG signals based on light emitted by flash and received by a camera (Jonathan and Leahy 2010, Grimaldi *et al* 2011). Smartphones are interesting devices in ambulatory scenarios due to significant advancements in the computational power which enables complex signal processing algorithms to be performed in real time. Certainly, built-in wireless communications feature of the smartphones facilitates ease of data transfer. These features make smartphones very valuable as ‘take-anywhere’ and easy-to-use physiological monitors (Scully *et al* 2012). Obtaining respiratory rates from smartphone devices would represent a simple and automated way for assisting hospital clinical staff who are currently trained to measure it by counting the number of breaths in a 15 or 30 s window (Pimentel *et al* 2014), making the process cumbersome and user-dependent. Other potential applications may include anxiety, fatigue or stress monitoring at home as respiratory rate is known to change in different anxiety/fatigue/stress situations (Marcora *et al* 2008, Niccolai *et al* 2009, Lackner *et al* 2011, Martinez *et al* 2015), especially if respiratory rate information is combined with other physiological information accessible in the PPG signal, such as pulse rate and its variability (Gil *et al* 2010) or blood pressure (Shaltis *et al* 2006).

It should be noted that, however, smartphone-camera-acquired-PPG (SCPPG) signal is more vulnerable to ambient-light interferences and variations in finger pressure over the sensor, making them in general noisier than the standard pulse oximeter sensor. Furthermore, their sampling rate is lower. Thus, deriving physiological information from SCPPG signals remains a more challenging situation than deriving it from conventional PPG signals, and the performance of known methods which have been tested with conventional PPG signals must be tested also with SCPPG signals.

In this paper, some PPG-based methods for deriving respiratory rate are studied with SCPPG signals. Concretely, the methods based on PRV, PAV, and PWV presented in (Lázaro *et al* 2013) are adapted to SCPPG signals. Furthermore, these methods are also combined with the AM- and FM-based methods presented in (Chon *et al* 2009). To the best of our knowledge, the PRV, PAV, and PWV-based methods have never been applied to SCPPG signals. In contrast, the AM- and FM-based methods have been tested with SCPPG signals in previous works (Scully *et al* 2012, Nam *et al* 2014). However, AM- and FM-based methods were neither combined with each other, nor with other methods. Note that a preliminary stage of the study described in this paper has been previously presented as a short conference paper (Lázaro *et al* 2014c). The present study is more comprehensive and new data are presented in this paper, including the study of several smartphone models with different hardware and form factor, which is relevant from the point of view of SCPPG signal acquisition.

2. Materials and methods

2.1. Data and signal preprocessing

We collected SCPPG data from 30 healthy subjects (22 men and eight women, between 20 and 26 years old) during controlled respiration experiments. Subjects were instructed to breathe at a constant rate according to a timed beeping sound, while sitting on a chair and placing the right index finger on the camera lens of the analyzed device. The data were collected for respiratory rates ranging from 0.2 to 0.6 Hz at an increment of 0.1 Hz, recording a total of 2 min of SCPPG signal for each subject, respiratory rate and device.

The SCPPG signals were recorded with 3 different smartphone devices: iPhone 4S, iPod 5, and HTC One M8. The signals were extracted from average of 50×50 pixel region of the green video signal at each frame. The reason for using only the green band is that there is high absorption by hemoglobin in the green range, and it has been demonstrated to give a stronger cardiac pulse signal than the red or blue bands during remote PPG imaging (Verkrusse *et al* 2008, Maeda *et al* 2011, Scully *et al* 2012, Matsumura *et al* 2014).

The sampling rate of SCPPG signals is variable due to internal processing load (Lee *et al* 2012), and it depends on the measuring device. The SCPPG signals were interpolated to a constant sampling rate of $f_s = 100$ Hz by using cubic splines. Furthermore, SCPPG signals are obtained as inverted PPG signals (Grimaldi *et al* 2011). Thus, the signals were inverted by multiplying by -1 to be used for further processing.

Next, the data were divided into 60 s-length data segments that were shifted every 10 s. A length of 60 s ensure at least 9 breaths of the lowest frequency eligible as respiratory rate in this work, which is 0.15 Hz. The baseline contamination was removed with a high-pass filter with a cutoff frequency of 0.3 Hz, and high frequency noise was considerably attenuated by a low-pass filter with a cutoff frequency of 35 Hz. Subsequently, the artifacts were automatically detected and removed by an algorithm based on Hjorth parameters described in (Gil *et al* 2008). Segments with 30% or more of the time containing artifactual signal were discarded.

2.2. Pulse-to-pulse methods

2.2.1. Significant points detection. SCPPG pulses apex points n_{A_i} were detected by an automatic PPG pulse detector based on a low-pass-differentiator filter and a time-varying threshold (Lázaro *et al* 2014b). Then, baseline point of the i th SCPPG pulse n_{B_i} was defined as the minimum previous to n_{A_i} :

$$n_{B_i} = \underset{n}{\operatorname{argmin}}\{x(n)\}, \quad n \in [n_{A_i} - 0.3f_s, n_{A_i}], \quad (1)$$

where $x(n)$ denotes the SCPPG signal.

Another significant point of SCPPG pulses is the middle point n_{M_i} . It is defined as the point where $x(n)$ has reached half of the maximum pulse amplitude, as shown in equation (2). These n_{M_i} were taken as fiducial points for deriving the pulse rate, because they are located at the upslopes of the SCPPG pulses which represent a very abrupt zone of $x(n)$, so their location is robust against noise (Lázaro *et al* 2014b).

$$n_{M_i} = \underset{n}{\operatorname{argmin}}\left\{\left|x(n) - \frac{x(n_{A_i}) + x(n_{B_i})}{2}\right|\right\}, \quad n \in [n_{B_i}, n_{A_i}]. \quad (2)$$

The SCPPG pulses width was measured by adapting the algorithm presented in (Lázaro *et al* 2013) for conventional PPG signals, based on pulse boundaries detection. Each pulse has two

boundaries: the onset n_{O_i} and the end n_{E_i} . They are detected by using a low-pass filtered first derivative of $x(n)$:

$$x'(n) = x_{LP}(n) - x_{LP}(n-1), \quad (3)$$

where $x_{LP}(n)$ is the low-pass filtered version of $x(n)$, using a cutoff frequency of f_c which was set to 2 Hz as shown in section 3.1.

The maximum upslope point n_{U_i} is defined as:

$$n_{U_i} = \underset{n}{\operatorname{argmax}} \{x'(n)\}, \quad n \in [n_{A_i} - 0.4f_s, n_{A_i}]. \quad (4)$$

Note that the interval for searching n_{U_i} is larger than in (Lázaro *et al* 2013), where its length was 300 ms. This is done because SCPPG signals are reflected-light-based signals so their pulses are smoother and larger than those in transmitted-light-based PPG signals. Similarly, the interval in which the search for pulse wave onset n_{O_i} is also larger:

$$n_{O_i} = \underset{n}{\operatorname{argmin}} \{|x'(n) - \eta x'(n_{U_i})|\}, \quad n \in \Omega_{O_i} = [n_{A_i} - 0.4f_s, n_{U_i}], \quad (5)$$

where $\eta x'(n_{U_i})$ represents a pulse-to-pulse varying threshold dependent on maximum upslope value of each pulse wave. The value of parameter η was set to 0.5 as shown in section 3.1.

Detection of pulse waves ends n_{E_i} was performed in a similar way as n_{O_i} but using maximum downslope n_{D_i} instead of n_{U_i} and $\Omega_{E_i} = [n_{D_i}, n_{A_i} + 0.4f_s]$ instead of Ω_{O_i} .

2.2.2. Derived respiration signals. Three derived respiration (DR) signals were calculated by using pulse-to-pulse methods: PRV, PAV, and PWV. The DR signal based on PRV is obtained through the inverse interval function (Sörnmo and Laguna 2005):

$$d_{PRV}^u(n) = \sum_i f_s \frac{1}{n_{N_i} - n_{N_{i-1}}} \delta(n - n_{A_i}), \quad (6)$$

where superscript ‘ u ’ denotes that the signal is unevenly sampled, and n_{N_i} are the arrival times of normal sinus pulses, which are determined from n_{M_i} after removing ectopic and missed-detected pulses using the method proposed in (Mateo and Laguna 2003).

On the other hand, the PAV- and PWV-based DR signals are defined as:

$$d_{PAV}^u(n) = \sum_i [x(n_{A_i}) - x(n_{B_i})] \delta(n - n_{A_i}) \quad (7)$$

$$d_{PWV}^u(n) = \sum_i \frac{1}{f_s} [n_{E_i} - n_{O_i}] \delta(n - n_{A_i}). \quad (8)$$

A median absolute deviation (MAD)-based outlier rejection rule described in (Bailón *et al* 2006a), and a 4 Hz evenly sampled version of each DR signal by cubic spline interpolation were applied. Then, these evenly sampled signals were filtered with a band-pass filter (0.15–0.7 Hz). The resulting signals are denoted without the superscript ‘ u ’, e.g. $d_{PWV}(n)$ is the outlier-rejected, evenly sampled, band-pass filtered version of $d_{PWV}^u(n)$.

Figure 1 illustrates these pulse-to-pulse derived respiration signals.

2.3. Non-pulse-to-pulse methods

Amplitude and frequency modulation sequences, $d_{AM}(n)$ and $d_{FM}(n)$ respectively, were extracted from SCPPG signal as described in (Chon *et al* 2009). The amplitude and frequency

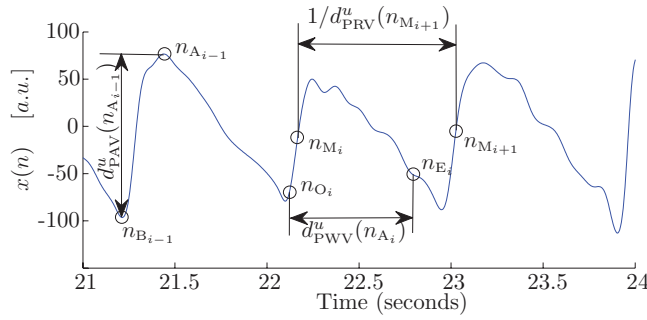


Figure 1. Pulse-to-pulse based derived respiration signals.

modulation sequences are extracted from a time-frequency (TF) spectrum obtained by the variable frequency complex demodulation (VFCDM) method (Wang *et al* 2006). The method for obtaining the VFCDM-based time-frequency spectrum can be divided into 2 steps: estimation of the dominant frequencies by fixed frequency complex demodulation (FFCDM), and subsequently applying VFCDM selecting only those dominant frequencies in order to improve the time-frequency resolution.

2.3.1. Fixed frequency complex demodulation. Let $x(t)$ be a narrow-band oscillation:

$$x(t) = x_{DC}(t) + A(t) \cos(2\pi f_0 t + \phi(t)), \quad (9)$$

where f_0 is the center frequency, $A(t)$ is the instantaneous amplitude, $\phi(t)$ is the phase and $x_{DC}(t)$ is the dc component.

$A(t)$ and $\phi(t)$ can be extracted for a given f_0 from $x(t)$ by shifting f_0 to zero frequency multiplying it by $e^{-j2\pi f_0 t}$:

$$z(t) = x_{DC}(t)e^{-j2\pi f_0 t} + \frac{A(t)}{2}e^{-j\phi(t)} + \frac{A(t)}{2}e^{-j(4\pi f_0 t + \phi(t))}. \quad (10)$$

Then, the middle term of (10) can be obtained from $z(t)$, by applying a low-pass filter with a cutoff-frequency lower than f_0 :

$$z_{LP}(t) = \frac{A(t)}{2}e^{-j\phi(t)}, \quad (11)$$

from which $A(t)$ and $\phi(t)$ can be obtained as:

$$A(t) = 2|z_{LP}(t)| \quad (12)$$

$$\phi(t) = \tan^{-1} \left(\frac{\text{Im}(z_{LP}(t))}{\text{Re}(z_{LP}(t))} \right). \quad (13)$$

2.3.2. Variable frequency complex demodulation. Consider now that the modulating frequency varies as a function of time, $f_0(t)$. equation (9) can be rewritten as:

$$x(t) = x_{DC}(t) + A(t) \cos \left(\int_0^t 2\pi f_0(\tau) d\tau + \phi(t) \right) \quad (14)$$

The frequency shift in (10) can be performed this time by multiplying $x(t)$ by $e^{-j \int_0^t 2\pi f_0(\tau) d\tau}$ obtaining:

$$z(t) = x_{\text{DC}}(t)e^{-j\int_0^t 2\pi f_0(\tau)d\tau} + \frac{A(t)}{2}e^{-j\phi(t)} + \frac{A(t)}{2}e^{-j\left(\int_0^t 4\pi f_0(\tau)d\tau + \phi(t)\right)}, \quad (15)$$

from which the middle term can be obtained similarly to the FFCDM case, i.e. by using a low-pass filter with a cut-off frequency lower than $f_0(t)$. Note that the expression of this term is the same than in the FFCDM case in (11) and thus, $A(t)$ and $\phi(t)$ can be obtained in the same way (see equations (12) and (13)). Then, the instantaneous frequency can be obtained as:

$$f(t) = f_0(t) + \frac{1}{2\pi} \frac{d\phi(t)}{dt}. \quad (16)$$

In this way, a time-frequency spectrum can be obtained by first applying FFCDM using a set of frequencies:

$$f_{0k} = (k-1)(2f_\omega), \quad k = 1, 2, \dots, \text{int}\left(\frac{f_{\max}}{2f_\omega}\right). \quad (17)$$

where $2f_\omega$ is the bandwidth between successive center frequencies and $f_{\max}$ denotes the highest signal frequency.

The dominating frequencies $f_k(t)$ can be obtained from (16), and $A_k(t)$ can be obtained from (12). Subsequently, $f_k(t)$ were used as central frequencies for applying VFCDM refining the time-frequency resolution.

Parameter f_ω was set to 0.6 Hz. Further details are given in (Chon *et al* 2009).

2.3.3. Derived respiration signals. Once the VFCDM-based TF spectrum $S_{\text{VFCDM}}(n, f)$ is computed, $d_{\text{FM}}(n)$ is determined by extracting the frequency component that has the largest amplitude for each time point at the heart rate frequency band, since heart rate is considered the carrier wave:

$$d_{\text{FM}}(n) = \underset{f \in \Omega_{\text{HR}}}{\text{argmax}} \{S_{\text{VFCDM}}(n, f)\}, \quad (18)$$

where Ω_{HR} denotes the frequency band in which heart rate is expected. This band is defined by using the spectrum of the SCPPG signal $S_{\text{SCPPG}}(f)$:

$$f_{\text{HR}} = \underset{f}{\text{argmax}} \{S_{\text{SCPPG}}(f)\}, \quad f \in [0.5\text{Hz}, 2\text{Hz}] \quad (19)$$

$$\Omega_{\text{HR}} = [f_{\text{HR}} - 0.2\text{Hz}, f_{\text{HR}} + 0.3\text{Hz}]. \quad (20)$$

A similar procedure is used for extracting the amplitude modulation:

$$d_{\text{AM}}(n) = \underset{f}{\text{max}} \{S_{\text{VFCDM}}(n, f)\}, \quad f \in \Omega_{\text{HR}}. \quad (21)$$

The values for parameters in these non-pulse-to-pulse methods were studied in previous works (Chon *et al* 2009, Scully *et al* 2012). The same processing applied to pulse-to-pulse-methods-based DR signals (section 2.2.2) was applied also to $d_{\text{FM}}(n)$ and $d_{\text{AM}}(n)$, i.e. a 4 Hz cubic spline interpolation followed by a band-pass filter (0.15–0.7 Hz). Figure 2 shows an example of DR signals studied in this paper.

2.4. Respiratory rate estimation

The respiratory rate is estimated from DR signals by an adaptation of the algorithm presented in (Lázaro *et al* 2013). It can combine information from several DR signals increasing the

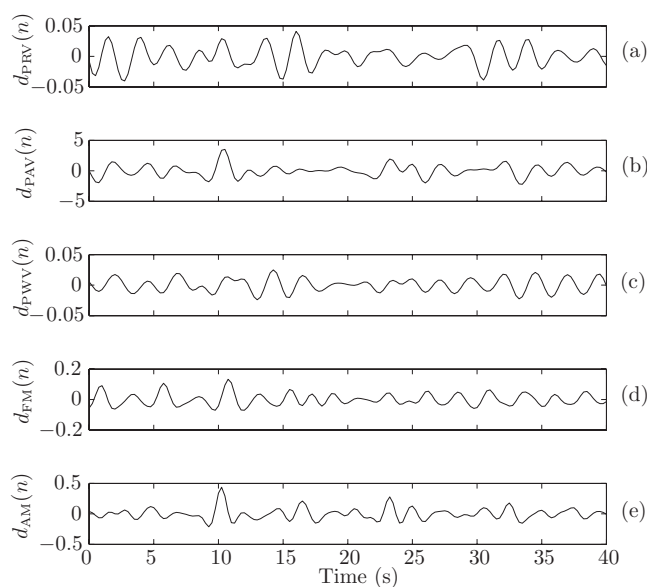


Figure 2. Example of derived respiration (DR) signals studied in this paper: $d_{PRV}(n)$ (a), $d_{PAV}(n)$ (b), $d_{PWV}(n)$ (c), $d_{FM}(n)$ (d), and $d_{AM}(n)$ (e). In this example, the subject was asked to maintain a respiratory rate of 0.4 Hz.

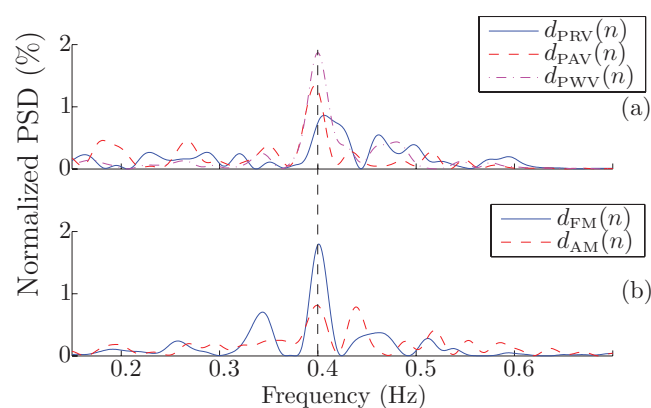


Figure 3. Example of normalized power spectrum densities (PSD) of derived respiration signals studied in this paper: $d_{PRV}(n)$, $d_{PAV}(n)$ and $d_{PWV}(n)$ (a), and $d_{FM}(n)$ and $d_{AM}(n)$ (b). In this example, the subject was asked to maintain a respiratory rate of 0.4 Hz.

robustness of the estimation. The algorithm can be divided in two phases: power spectrum density (PSD) estimation, and respiratory rate estimation.

First, the PSD of the j th DR signal $S_j(f)$ was estimated by applying a modified periodogram using a Hamming window. Then, the respiratory rate $\hat{f}$ is the frequency at where the absolute maximum of the PSD is located, within the studied band [0.15, 0.7 Hz]. Figure 3 shows an example of normalized PSDs of DR signals studied in this paper.

Because $\hat{f}$ is being estimated from more than one DR signals, their PSDs are ‘peaked-condition averaged’; only those $S_j(f)$ which are sufficiently peaked take part in the averaging.

In this paper, ‘peaked’ denotes that a certain percentage (ξ) of PSD must be contained in an interval around its highest peak. In mathematical terms, ‘peakness’ of a PSD is defined as:

$$P_j = \frac{\int_{f_p(j)-0.05\text{Hz}}^{f_p(j)+0.05\text{Hz}} S_j(f)df}{\int_{0.15\text{Hz}}^{0.7\text{Hz}} S_j(f)df}, \quad (22)$$

where $f_p(j)$ denotes the highest peak within the studied band [0.15, 0.7 Hz] in the PSD of the j th DR signal.

In order to select those spectra that are sufficiently ‘peaked’, two different criteria were established: χ^A and χ^B . On the one hand, χ^A lets those spectra whose ‘peakness’ is greater than a fixed value take part in the average as shown in equation (14). On the other hand, χ^B compares the spectra of different DR signals, letting those spectra more peaked take part in the average, although all of them have passed the χ^B criterion as shown in equation (15).

$$\chi_j^A = \begin{cases} 1, & P_j \geq \xi \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

$$\chi_j^B = \begin{cases} 1, & P_j \geq \max_j \{P_j\} - \lambda \\ 0, & \text{otherwise} \end{cases}. \quad (24)$$

Then, the ‘peak-conditioned’ average is computed as:

$$\bar{S}(f) = \sum_j \chi_j^A \chi_j^B S_j(f). \quad (25)$$

Finally, $\hat{f}$ is estimated as the frequency at which the absolute maximum of $\bar{S}(f)$ is located within the studied band [0.15, 0.7 Hz]:

$$\hat{f} = \underset{f \in [0.15, 0.7]}{\operatorname{argmax}} \{ \bar{S}(f) \}. \quad (26)$$

Respiratory rate was estimated from each one of the five DR signals separately, and from two combinations:

- $C_{\text{PRV,PAV,PWV}}$: $d_{\text{PRV}}(n)$, $d_{\text{PAV}}(n)$ and $d_{\text{PWV}}(n)$
- C_{ALL} : $d_{\text{PRV}}(n)$, $d_{\text{PAV}}(n)$, $d_{\text{PWV}}(n)$, $d_{\text{AM}}(n)$ and $d_{\text{FM}}(n)$

3. Results

3.1. Pulse width parameters optimization

Optimal values for f_c and η parameters of pulse width measurement algorithm were obtained using a similar procedure to that used in (Lázaro *et al* 2013). Respiratory rate estimates from $d_{\text{PWV}}(n)$ were computed for all the 323 possible combinations corresponding to $\eta \in [0, 0.8]$ with a step of 0.05, and $f_c \in [1, 10]$ Hz with a step of 0.5 Hz. The relative error of estimated respiratory rate was obtained as:

$$e_R = \frac{\hat{f} - f_R}{f_R} \times 100, \quad (27)$$

Table 1. Percentage of fragments excluded from the study due to the artifact and aliasing criteria.

f_R (Hz)	iPhone 4S		iPod 5		HTC One M8	
	Artifact	Aliasing	Artifact	Aliasing	Artifact	Aliasing
0.2	15.38%	0.00%	8.47%	0.00%	20.98%	0.00%
0.3	17.51%	0.00%	9.63%	0.00%	19.52%	0.00%
0.4	16.36%	0.00%	6.95%	0.00%	18.10%	0.00%
0.5	14.03%	1.36%	4.79%	6.38%	22.13%	0.00%
0.6	14.09%	24.55%	7.53%	35.48%	8.57%	15.71%

where f_R denotes the rate at which subject is requested to breathe.

Then, the values that minimized the mean of absolute value of e_R were obtained and chosen as optimal. These values were the same for the 3 studied devices: $\eta = 0.5$ and $f_c = 2$ Hz.

3.2. Respiratory rate estimation

The percentage of 60 s-length fragments excluded by the artifact criterion described in section 2.1 is shown in table 1. Note that aliasing problems may affect pulse-to-pulse methods, since respiratory information is obtained only at pulse occurrence. For this reason, fragments associated to a respiratory rate higher than the half mean pulse rate were excluded from the study. The percentage of fragments excluded by this criterion is also shown in table 1.

Relative error e_R was obtained for each studied DR signal and combination as defined in equation (27). Medians and interquartile ranges (IQR) obtained for e_R from different DR signals and combinations, for each f_R and device, are shown in table 2, and figure 4 shows them in a boxplot for a graphical visualization.

Furthermore, Kruskal-Wallis and the Bonferroni t test were used for analysis of differences of e_R for the different methods. The non-parametric Kruskal-Wallis statistical test was chosen because it was observed that e_R is not normal distributed, and the Bonferroni correction was applied in order to control the familywise error rate because multiple comparisons were performed. Table 3 shows those methods for which significant differences (p -value < 0.05) were observed.

4. Discussion

In this paper, two methods for deriving respiratory rate from SCPPG signals are presented. One of them combines information from pulse-to-pulse methods PRV, PAV and PWV, which were previously studied with conventional pulse oximeter PPG signals (Lázaro *et al* 2013). The other method presented in this paper uses the pulse-to-pulse methods in combination with non-pulse-to-pulse methods presented in (Chon *et al* 2009).

Deriving information from SCPPG signals is one challenging issue, since their low sampling rate and the ambient-light noise considerably affect their quality. In order to deal with this issue, an artifact detector (Gil *et al* 2008) was used to automatically exclude the artifactual fragments, which represents up to a 22.13% of the total fragments (HTC at $f_R = 0.5$ Hz).

The metronome frequency was used as reference for respiratory rate because a respiratory signal was not available when data were collected for the iPod 5 and HTC One M8 experiments. However, a respiratory signal from a respiration belt was available for 10 subjects in the iPhone 4S experiments and according to it, subjects breathed at the metronome respiratory

Table 2. Obtained medians and interquartile ranges (IQR) for e_R from different derived respiration signals and combinations, for each f_R and device.

f_R (Hz)	DR Signal/Combination	iPhone 4S		iPod 5		HTC One M8	
		Median	IQR	Median	IQR	Median	IQR
0.2	$d_{FM}(n)$	0.10%	2.44%	0.10%	0.00%	0.10%	2.44%
	$d_{AM}(n)$	0.10%	2.44%	0.10%	4.88%	0.10%	4.88%
	$d_{PRV}(n)$	0.10%	1.95%	0.10%	1.46%	0.10%	0.98%
	$d_{PAV}(n)$	0.10%	2.93%	0.10%	3.05%	1.07%	15.87%
	$d_{PWV}(n)$	0.10%	1.46%	0.10%	1.46%	0.10%	1.46%
	$C_{PRV,PAV,PWV}$	−0.39%	1.46%	−0.39%	0.98%	−0.39%	0.98%
	C_{ALL}	−0.39%	1.10%	−0.39%	0.98%	−0.39%	0.98%
0.3	$d_{FM}(n)$	0.91%	1.63%	−0.72%	1.63%	0.91%	18.31%
	$d_{AM}(n)$	−0.72%	3.66%	−0.72%	2.03%	−0.72%	3.26%
	$d_{PRV}(n)$	−0.07%	0.98%	−0.07%	1.06%	−0.07%	1.38%
	$d_{PAV}(n)$	−0.07%	1.95%	−0.39%	1.38%	−0.39%	2.36%
	$d_{PWV}(n)$	0.10%	0.98%	−0.07%	0.65%	−0.39%	1.63%
	$C_{PRV,PAV,PWV}$	0.91%	0.65%	−0.07%	0.98%	−0.07%	0.73%
	C_{ALL}	−0.07%	0.65%	−0.07%	0.98%	−0.07%	0.98%
0.4	$d_{FM}(n)$	0.10%	4.88%	0.10%	1.22%	0.10%	10.99%
	$d_{AM}(n)$	−2.34%	41.50%	−1.12%	36.62%	−1.12%	34.18%
	$d_{PRV}(n)$	−0.15%	1.22%	−0.15%	1.22%	0.10%	2.20%
	$d_{PAV}(n)$	−0.15%	3.17%	−0.39%	24.66%	−0.63%	15.38%
	$d_{PWV}(n)$	−0.15%	1.46%	−0.15%	1.22%	−0.15%	2.69%
	$C_{PRV,PAV,PWV}$	−0.15%	0.49%	−0.15%	0.73%	−0.15%	1.22%
	C_{ALL}	−0.15%	0.73%	−0.15%	0.73%	−0.15%	0.98%
0.5	$d_{FM}(n)$	−0.39%	3.17%	−0.39%	1.95%	−0.39%	11.72%
	$d_{AM}(n)$	−39.94%	55.91%	−39.45%	59.57%	−25.78%	56.64%
	$d_{PRV}(n)$	−0.20%	4.64%	−0.20%	1.17%	−0.20%	11.91%
	$d_{PAV}(n)$	−0.59%	44.14%	−0.39%	45.90%	−0.20%	26.56%
	$d_{PWV}(n)$	−0.20%	2.29%	0.00%	1.95%	−1.37%	39.06%
	$C_{PRV,PAV,PWV}$	−0.20%	0.78%	0.00%	0.78%	−0.20%	1.17%
	C_{ALL}	0.00%	0.98%	0.00%	0.78%	−0.20%	4.69%
0.6	$d_{FM}(n)$	−0.72%	50.05%	0.10%	46.39%	−3.97%	44.76%
	$d_{AM}(n)$	−49.95%	65.10%	−57.28%	66.73%	−37.74%	59.41%
	$d_{PRV}(n)$	−0.47%	36.42%	−0.23%	4.11%	−13.90%	56.32%
	$d_{PAV}(n)$	−2.99%	59.57%	−51.66%	69.42%	−31.64%	54.32%
	$d_{PWV}(n)$	−0.39%	32.63%	−0.07%	8.63%	−14.71%	48.50%
	$C_{PRV,PAV,PWV}$	−0.23%	2.40%	−0.07%	3.01%	−13.49%	54.57%
	C_{ALL}	−0.23%	2.12%	−0.07%	1.99%	−2.51%	37.64%

rate with an error of 0.12/1.01 mHz (median/interquartile range) which is accurate enough to consider the metronome frequency as a reference.

A high-pass filter with a cut-off frequency of 0.3 Hz was applied to SCPPG signals in order to significantly attenuate the baseline. Although in some situations respiration is below 0.3 Hz, this filter does not attenuate the respiration-induced variations in the amplitude of SCPPG signal exploited by some of the studied methods (PAV and AM). On one hand, PAV is based on pulse amplitude with respect to the baseline (see equation (7)). In this way, the information

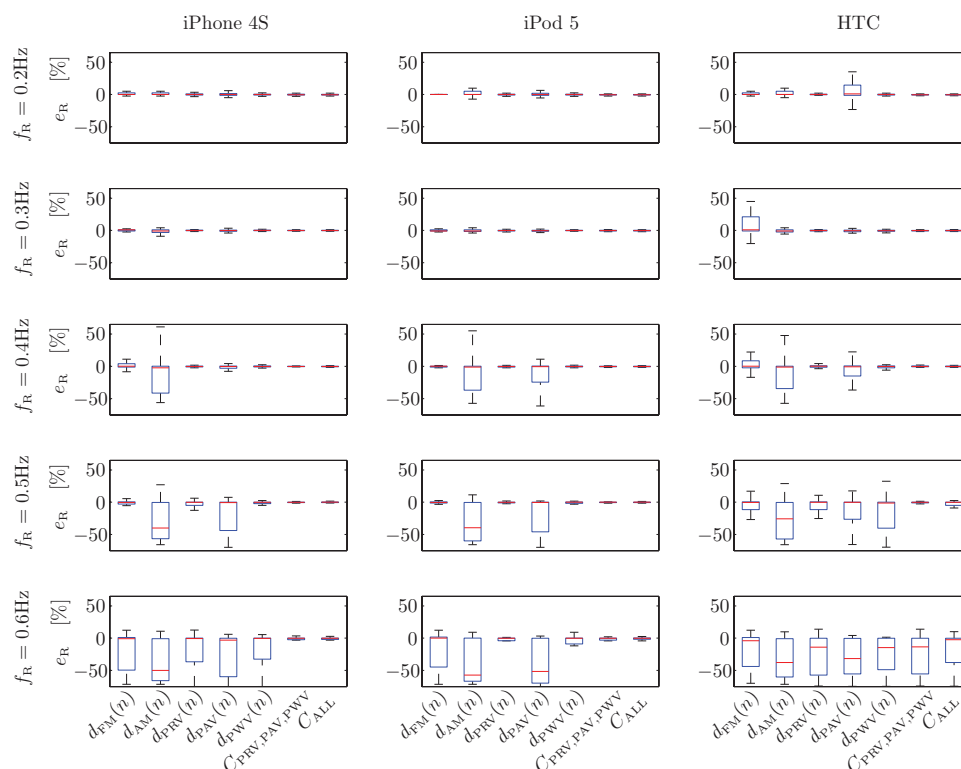


Figure 4. Boxplots of relative error e_R for the different methods, devices, and respiratory rates f_R .

Table 3. Pairs of methods for which significant differences were found in obtained e_R for different studied devices. Obtained e_R for normal range of spontaneous respiratory rate (0.2, 0.3 and 0.4 Hz) and for higher respiratory rates (0.5 and 0.6 Hz) were studied separately.

	$f_R \in \{0.2, 0.3, 0.4\}$ Hz	$f_R \in \{0.5, 0.6\}$ Hz
iPhone 4S	{FM, AM}, {FM, PRV}, {FM, PAV}, {FM, PWV}, {FM, C _{PRV,PAV,PWV} }, {FM, C _{ALL} }, {AM, PWV}	{FM, AM}, {FM, PRV}, {AM, PRV}, {AM, PAV}, {AM, PWV}, {AM, C _{PRV,PAV,PWV} }, {AM, C _{ALL} }, {PRV, PAV}, {PAV, PWV}, {PAV, C _{PRV,PAV,PWV} }, {PAV, C _{ALL} }
iPod 5	{FM, AM}, {FM, PAV}, {AM, PRV}, {AM, PAV}, {AM, PWV}, {AM, C _{PRV,PAV,PWV} }, {AM, C _{ALL} }, {PRV, PAV}, {PAV, PWV}, {PAV, C _{PRV,PAV,PWV} }, {PAV, C _{ALL} }	{FM, AM}, {FM, PAV}, {AM, PRV}, {AM, PAV}, {AM, PWV}, {AM, C _{PRV,PAV,PWV} }, {AM, C _{ALL} }, {PRV, PAV}, {PAV, PWV}, {PAV, C _{PRV,PAV,PWV} }, {PAV, C _{ALL} }
HTC	{FM, AM}, {FM, PRV}, {FM, PAV}, {FM, PWV}, {FM, C _{PRV,PAV,PWV} }, {FM, C _{ALL} }	{FM, AM}, {FM, PAV}, {FM, PWV}, {AM, PRV}, {AM, C _{PRV,PAV,PWV} }, {AM-C _{ALL} }

exploited by the PAV-based method is in the upwards slope of the pulses which correspond to a higher frequency. On the other hand, the AM method performs an amplitude demodulation considering the pulse rate to be the hypothetical carrier. Thus, the lower frequency exploited by the AM-based method is the pulse rate minus the respiratory rate, which is over 0.3 Hz.

In general, all studied methods obtained low median (of order 0.5%) and low IQR (of order 2.5%) of e_R until reaching a given respiratory rate, which depends on the method and on the device, e.g. $d_{PWV}(n)$ maintain good performance in e_R terms up to 0.5 Hz when using the iPhone 4S, and up to 0.4 Hz when using the HTC One M8. Similarly, in mathematical terms, no method was especially disadvantaged at higher respiratory rates. A possible reason for this observation is that the respiration-induced modulations on which DR signals are based (rate, amplitude and width) may have a less strong effect at higher respiratory rates. In the case of pulse rate, it is known that respiratory sinus arrhythmia (which modulates the heart rate and therefore the pulse rate) is reduced at high respiratory rates.

Results obtained for $d_{PWV}(n)$ were comparable to those obtained for $d_{PRV}(n)$ and better than for the other DR signals in general, obtaining low medians and IQR for e_R with f_R up to 0.4 Hz and even 0.5 Hz when using the iPhone 4S and iPod 5 devices. Occasionally, $d_{PWV}(n)$ and $d_{PRV}(n)$ obtained worse results (higher median/IQR for e_R) than another DR signal, such as $d_{FM}(n)$ when using the iPod 5 device with $f_R = 0.2$ Hz (0.10/0.00% versus -0.10/1.46%), or when using the HTC One M8 device with $f_R = 0.4$ Hz (-0.39/11.72% versus -0.20/11.91% and -1.37/39.06%).

Both combinations $C_{PRV,PAV,PWV}$ and C_{ALL} obtained low median (less than 0.5%) and IQRs (less than 2.5%) for e_R , in every case where at least one of the DR signals included in the combinations obtained low median and IQR for e_R , and even in some cases where none of the DR signals obtained low median and IQR for e_R . For instance, in the case of the iPhone 4S at $f_R = 0.6$ Hz, combinations obtained low median and IQR for e_R although DR signals obtained very high IQRs for e_R (up to 65.10%). Similarly, at $f_R = 0.5$ Hz, both combinations still obtained low median and IQR for e_R , even though in this case $d_{AM}(n)$ and $d_{PAV}(n)$ obtained high IQR (55.91% and 44.14%, respectively). These observations demonstrate the advantages of combining information.

C_{ALL} obtained similar results to $C_{PRV,PAV,PWV}$ in e_R terms, and significant statistical differences were not found between their associated e_R , for all devices either at normal ranges of spontaneous respiratory rate (0.2–0.4 Hz) or higher ones (0.5–0.6 Hz). These results suggest that C_{ALL} offers no advantages over $C_{PRV,PAV,PWV}$. A possible reason for this may be that respiratory information in $d_{AM}(n)$ and $d_{FM}(n)$ is mainly redundant with respiratory information in $d_{PAV}(n)$, $d_{PRV}(n)$ and/or $d_{PWV}(n)$. It is reasonable to believe that respiratory information in $d_{AM}(n)$ and $d_{PAV}(n)$ is redundant to a large extent, because they are based on similar effects: respiration-induced amplitude modulations, of the SCPPG signal in one case, and of pulses of SCPPG in the other one. A similar case occurs with $d_{FM}(n)$ and $d_{PRV}(n)$. Note that statistical differences were found in some cases between $d_{AM}(n)$ and $d_{PAV}(n)$ (iPhone 4S at 0.5–0.6 Hz and iPod 5 at both 0.2–0.4 Hz and 0.5–0.6 Hz) and between $d_{FM}(n)$ and $d_{PRV}(n)$ (iPhone 4S at both 0.2–0.4 Hz and 0.5–0.6 Hz, and HTC One M8 at 0.2–0.4 Hz). However before interpreting this observation it must be kept in mind that when a method fails in tracking respiration, the obtained e_R , especially when errors are big, has clear tendencies, see figure 4, so statistical differences in e_R should not be considered as an indicator of differences in the physiological origin of those respiratory-related modulations. When the statistical tests is repeated taking only those e_R between -15 and 15% (excluding outliers), no statistical differences are found between these methods in any device/respiratory rate condition, so corroborating results independence with the used methodology in respiratory frequency derivation when the methods

are able to catch the respiration. The differences are then in the different ability to provide meaningful estimation.

PPG-amplitude- and rate-based derived respiration signals present low frequency modulations below 0.15 Hz due to the Mayer wave related to sympathetic activity, which can be considered as noise from the point of view of deriving respiratory information. The power of these modulations is usually comparable or even higher than the respiration-related modulations, and this may confound respiratory rate estimation. Regarding the methods studied in this work, the low-frequency modulation affect to the AM and FM methods (Chon *et al* 2009) and to the PAV and PRV methods, but not to the PWV method according to (Lázaro *et al* 2013). For this reason, results for respiratory rates below 0.15 Hz are not provided in this work, in order to study if respiration-related modulations already studied in conventional PPG signals are also present in SCPPG signals, with independence of this kind of noise.

Those fragments associated with a f_R higher than half mean pulse rate were excluded, because the pulse-to-pulse methods would track an alias in such situations. This problem affects high f_R (0.5 and 0.6 Hz); e.g. for tracking a $f_R = 0.6$ Hz using pulse-to-pulse methods, it would be necessary to have a mean pulse rate of 1.2 Hz, i.e. 72 beats per min. However, a high f_R with a low pulse rate does not represent a realistic physiological situation. In such situations when the autonomic nervous system requires a high respiratory rate, it also requires a high heart rate which leads to a high pulse rate, e.g. during exercise. Nevertheless, some medications affect autonomic nervous system and may lead to non-physiological situations with a high f_R with a low pulse rate, e.g. beta-blockers. Furthermore, the physiological source of the respiration-related modulations in SCPPG signal exploited by the presented methods is the autonomic control over the cardiovascular system. In addition, PPG pulses morphology is affected by age, due to arterial stiffness. So age, arterial or autonomic nervous system diseases or medications interactions could affect results. This remains one limitation of this study because the methods have been evaluated only with recordings from healthy young people. Further studies must be elaborated to assess the performance of the presented methods over this kind of patients.

Another limitation of this study is that the inter-device variability for the same model of smartphones cannot be assessed because only one device per model has been tested. Slight differences in flashlight or camera lens may affect results. However, the form factor and so the distance between flashlight and camera lens, which is the most important signal-acquisition difference between different smartphone models, remains the same for devices of the same model. Nevertheless, if different models of smartphone would be wanted to be compared in the task of deriving respiratory rate, further studies using several devices for each model should be elaborated.

5. Conclusions

Results suggest that normal ranges of spontaneous respiratory rates (0.2–0.4 Hz) can be accurately estimated from smartphone-camera-acquired pulse photoplethysmographic signals based on pulse width variability or pulse rate variability with low e_R (median on the order of 0.5% and IQR on the order of 2.5%). The accuracy can be further improved by combining them with other methods such as pulse rate and amplitude variabilities, and amplitude and/or frequency modulations. Indeed, the combination of these methods resulted in lower e_R values within normal ranges of spontaneous respiratory rate, but with small degradation in its performance at higher rates (up to 0.5 Hz when using HTC One M8, and up to 0.6 Hz when using the iPhone 4S or iPod 5 devices).

These promising results suggest that accurate normal ranges of respiratory rates can be obtained from general people using only smartphones without using any external sensors. The methods could be extended to other models of smartphones or tablet devices, the only requirement is that these smartphones and tablets contain a video camera to image a fingertip pressed to it. As smartphones and tablets have become common, they meet the criteria of ready access and acceptance. Hence, our mobile phone/tablet approach has the potential to be widely-accepted by the general population and can facilitate the capability to measure some of the vital signs using only the subject's fingertip.

Acknowledgments

This work was supported in part by the US Army Medical Research and Materiel Command (USAMRMC) under Grant No. W81XWH-12-1-0541, by Universidad de Zaragoza under fellowship PIFUZ-2011-TEC-A-003, by Ministerio de Economía y Competitividad (MINECO), FEDER; under projects TIN2014-53567-R, TIN2013-42140-R, TEC2013-42140-R and FIS-PI12/00514, by CIBER de Bioingeniería, Biomateriales y Nanomedicina through Instituto de Salud Carlos III, and by Grupo Consolidado BSICoS (T96) from DGA (Aragón) and European Social Fund (EU). The computation was performed by the ICTS 'NANBIOSIS', more specifically by the High Performance Computing Unit of the CIBER in Bioengineering, Biomaterials & Nanomedicine (CIBER-BBN) at the University of Zaragoza.

References

- Bailón R, Sörnmo L and Laguna P 2006a A robust method for ECG-based estimation of the respiratory frequency during stress testing *IEEE Trans. Biomed. Eng.* **53** 1273–85
- Bailón R, Sörnmo L and Laguna P 2006b ECG-derived respiratory frequency estimation *Advanced Methods and Tools for ECG Data Analysis* ed G Clifford *et al* (Norwood, MA: Artech House Inc) pp 215–44
- Chon K H, Dash S and Ju K 2009 Estimation of respiratory rate from photoplethysmogram data using time-frequency spectral estimation *IEEE Trans. Biomed. Eng.* **56** 2054–63
- Chua C P and Heneghan C 2005 Pulse transit time-derived respiratory parameters and their variability across sleep stages *Proc. of the Ann. Int. Conf. IEEE Engineering Medical Biology Society* pp 6153–6
- De Meersman R E, Zion A S, Teitelbaum S, Weir J P, Lieberman J and Downey J 1996 Deriving respiration from pulse wave: a new signal-processing technique *Am. J. Physiol.* **270** H1672–5
- Garde A, Karlen W, Ansermino J M and Dumont G A 2014 Estimating respiratory and heart rates from the correlogram spectral density of the photoplethysmogram *PLoS ONE* **9** e86427
- Garde A, Karlen W, Dehkordi P, Ansermino J and Dumont G 2013 Empirical mode decomposition for respiratory and heart rate estimation from the photoplethysmogram *Comput. Cardiol.* **2013** 799–802
- Gil E, María Vergara J and Laguna P 2008 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** 267–77
- Gil E, Orini M, Bailón R, Vergara J M, Mainardi L and Laguna P 2010 Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions *Physiol. Meas.* **31** 1271–90
- Gravelyn T R and Weg J G 1980 Respiratory rate as an indicator of acute respiratory dysfunction *JAMA* **244** 1123–5
- Grimaldi D, Kurylyak Y, Lamonaca F and Nastro A 2011 Photoplethysmography detection by smartphone's videocamera *Proc. of the IEEE 6th Int. Conf. on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)* vol 1 pp 488–91

- Hirsch J A and Bishop B 1981 Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate *Am. J. Physiol.* **241** H620–9
- Johansson A and Oberg P A 1999 Estimation of respiratory volumes from the photoplethysmographic signal. Part I: experimental results *Med. Biol. Eng. Comput.* **37** 42–7
- Jonathan E and Leahy M 2010 Investigating a smartphone imaging unit for photoplethysmography *Physiol. Meas.* **31** N79–83
- Krieger B, Feinerman D, Zaron A and Bizousky F 1986 Continuous noninvasive monitoring of respiratory rate in critically ill patients *Chest* **90** 632–4
- Lackner H K, Papousek I, Batzel J J, Roessler A, Scharfetter H and Hinghofer-Szalkay H 2011 Phase synchronization of hemodynamic variables and respiration during mental challenge *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **79** 401–9
- Larsen V H, Christensen P H, Oxhøj H and Brask T 1984 Impedance pneumography for long-term monitoring of respiration during sleep in adult males *Clin. Physiol. Oxford England* **4** 333–42
- Lázaro J, Alcaine A, Romero D, Gil E, Laguna P, Pueyo E and Bailón R 2014a Electrocardiogram derived respiratory rate from QRS slopes and R-wave angle *Ann. Biomed. Eng.* **42** 2072–83
- Lázaro J, Gil E, Bailón R, Mincholé A and Laguna P 2013 Deriving respiration from photoplethysmographic pulse width *Med. Biol. Eng. Comput.* **51** 233–42
- Lázaro J, Gil E, Vergara J M and Laguna P 2014b Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children *IEEE J. Biomed. Health Inform.* **18** 240–6
- Lázaro J, Nam Y, Gil E, Laguna P and Chon K H 2014c Smartphone-camera-acquired pulse photoplethysmographic signal for deriving respiratory rate *Proc. of the 8th Conf. of the European Study Group on Cardiovascular Oscillations (ESGCO)* pp 121–2
- Lee J, Reyes B A, McManus D D, Mathias O and Chon K H 2012 Atrial fibrillation detection using a smart phone *Proc. Ann. Int. Conf. IEEE Engineering Medical Biology Society* pp 1177–80
- Maeda Y, Sekine M and Tamura T 2011 The advantages of wearable green reflected photoplethysmography *J. Med. Syst.* **35** 829–34
- Marcora S M, Bosio A and de Morree H M 2008 Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independently from metabolic stress *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294** R874–83
- Martinez J M, Garakani A, Aaronson C J and Gorman J M 2015 Heart rate and respiratory response to doxapram in patients with panic disorder *Psychiatry Res.* **227** 32–8
- Mason C L and Tarassenko L 2001 Quantitative assessment of respiratory derivation algorithms *Proc. of the 23rd Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society* vol 2 pp 1998–2001
- Mateo J and Laguna P 2003 Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal *IEEE Trans. Biomed. Eng.* **50** 334–43
- Matsumura K, Rolfe P, Lee J and Yamakoshi T 2014 iPhone 4s photoplethysmography: which light color yields the most accurate heart rate and normalized pulse volume using the iPhysio meter application in the presence of motion artifact? *PloS One* **9** e91205
- Meredith D J, Clifton D, Charlton P, Brooks J, Pugh C W and Tarassenko L 2012 Photoplethysmographic derivation of respiratory rate: a review of relevant physiology *J. Med. Eng. Technol.* **36** 1–7
- Nam Y, Lee J and Chon K H 2014 Respiratory rate estimation from the built-in cameras of smartphones and tablets *Ann. Biomed. Eng.* **42** 885–98
- Niccolai V, van Duinen M A and Griez E J 2009 Respiratory patterns in panic disorder reviewed: a focus on biological challenge tests *Acta Psychiatr. Scand.* **120** 167–77
- Peláez-Coca M D, Orini M, Lázaro J, Bailón R and Gil E 2013 Cross time-frequency analysis for combining information of several sources: application to estimation of spontaneous respiratory rate from photoplethysmography *Comput. Math. Methods Med.* **2013** 631978
- Pimentel M A F, Santos M D, Arteta C, Domingos J S, Maraci M A and Clifford G D 2014 Respiratory rate estimation from the oscillometric waveform obtained from a non-invasive cuff-based blood pressure device *Proc. Ann. Int. Conf. IEEE Engineering Medical Biology Society* pp 3821–4
- Scully C G, Lee J, Meyer J, Gorbach A M, Granquist-Fraser D, Mendelson Y and Chon K H 2012 Physiological parameter monitoring from optical recordings with a mobile phone *IEEE Trans. Biomed. Eng.* **59** 303–6
- Shaltis P A, Reisner A and Asada H H 2006 Wearable, cuff-less PPG-based blood pressure monitor with novel height sensor *Proc. Ann. Int. Conf. IEEE Engineering Medical Biology Society* pp 908–11

- Sörnmo L and Laguna P 2005 ECG signal processing: heart rate variability *Bioelectrical Signal Processing in Cardiac and Neurological Applications* (Amsterdam: Elsevier)
- Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996 Heart rate variability standards of measurement, physiological interpretation, and clinical use *Circulation* **93** 1043–65
- Verkruysse W, Svaasand L O and Nelson J S 2008 Remote plethysmographic imaging using ambient light *Opt. Express* **16** 21434–45
- Wang H, Siu K, Ju K and Chon K H 2006 A high resolution approach to estimating time-frequency spectra and their amplitudes *Ann. Biomed. Eng.* **34** 326–38

Article

Tidal Volume Estimation Using the Blanket Fractal Dimension of the Tracheal Sounds Acquired by Smartphone

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

Department of Biomedical Engineering, University of Connecticut, 260 Glenbrook Road, Storrs, CT 06269, USA; E-Mails: bersain.reyes@uconn.edu (B.A.R.); kchon@engr.uconn.edu (K.H.C.)

* Author to whom correspondence should be addressed; E-Mail: reljin@engr.uconn.edu; Tel.: +1-860-486-5838; Fax: +1-860-486-2500.

Academic Editor: Vittorio M.N. Passaro

Received: 5 March 2015 / Accepted: 21 April 2015 / Published: 27 April 2015

Abstract: In this paper, we propose the use of blanket fractal dimension (BFD) to estimate the tidal volume from smartphone-acquired tracheal sounds. We collected tracheal sounds with a Samsung Galaxy S4 smartphone, from five ($N=5$) healthy volunteers. Each volunteer performed the experiment six times; first to obtain linear and exponential fitting models, and then to fit new data onto the existing models. Thus, the total number of recordings was 30. The estimated volumes were compared to the true values, obtained with a RespiTrace system, which was considered as a reference. Since Shannon entropy (SE) is frequently used as a feature in tracheal sound analyses, we estimated the tidal volume from the same sounds by using SE as well. The evaluation of the performed estimation, using BFD and SE methods, was quantified by the normalized root-mean-squared error (NRMSE). The results show that the BFD outperformed the SE (at least twice smaller NRMSE was obtained). The smallest NRMSE error of $15.877\% \pm 9.246\%$ (mean $\pm$ standard deviation) was obtained with the BFD and exponential model. In addition, it was shown that the fitting curves calculated during the first day of experiments could be successfully used for at least the five following days.

Keywords: blanket fractal dimension; tidal volume; tracheal sounds; smartphone

1. Introduction

Tracheal sounds are defined as those that are detected or heard over the extrathoracic part of the trachea [1]. Tracheal sounds are strong, and cover a wide frequency range [2]. As part of respiratory sounds, they play an important role in monitoring respiratory activity, as well as in detection of pulmonary diseases [1–3].

Respiratory activity is one of the vital signs, and as such requires an adequate attention. Tidal volume is one of the parameters for monitoring respiratory activity [4]. It plays an important role for both healthy people and people with respiratory diseases, hence measuring and checking volume's values can be helpful, especially in assessing risky situations involving respiratory failure [4–6]. Tidal volume is defined as the volume of air exchanged in one breath, and is commonly measured at the mouth [1,2,7]. The average value is about 500 mL per breath at rest [2,7]. Various methods exist for measuring the tidal volume, such as spirometry, whole-body plethysmography, inductance plethysmography, and electrocardiography [2,8–10]. However, these methods require the use of specialized equipment, and cannot be easily applied in nonclinical settings. Therefore, there is a need for a miniature monitoring device that can be used in everyday situations and not only in clinical and/or research settings [11]. In addition, with an extensive growth of electronic devices and their computational capabilities, the development of portable tidal volume estimation systems is now possible [12].

Several efforts have been made in the research oriented towards the estimation of tidal volume. In [13], the authors estimated volume by optically tracking reflective markers in three dimensions. Petrovic *et al.* proposed a technique for measuring tidal volumes by using a single fiber-grating sensor [14], while in [15] the authors estimated the tidal volume using Doppler radar signals. Chen *et al.* estimated tidal volume from the energy of the tracheal sounds [6]. To the best of our knowledge, there are no studies exploring the possibility to estimate tidal volume directly from smartphone-acquired tracheal sounds.

Smartphones are widely used nowadays. They have fast microprocessors, large storage capacities and a lot of media capabilities. In addition, the mobility of the smartphones is making them more popular for usage outside the clinics or research facilities, when they can be used for measuring vital signs and health monitoring, as shown in some of the previous works of our research group [16–18].

In this paper, we propose the use of blanket fractal dimension (BFD) for estimating the tidal volume from tracheal sounds acquired by a commercially available Android smartphone. Tracheal sounds, as part of respiratory sounds, are non-stationary and stochastic signals [2,19]. Due to this fact, some past studies investigated and showed successful applications of fractal analysis on tracheal and lung sounds [20–24]. None of these efforts was concerned with the tidal volume estimation using fractal analysis. In this study, we explore the possibility to estimate tidal volume using BFD, which, to the best of our knowledge, was not used for respiratory sound analysis. The estimated volumes were compared to peak-to-peak volumes obtained from a Respitrace signal, which was considered as a reference. In addition, we estimated volumes by obtaining Shannon entropy (SE) from the same tracheal sounds, and compared them to reference volumes. For testing the proposed method and comparing it with SE method, we collected signals from healthy and non-smoker volunteers for six days, for a total of 30 recordings. As a figure of merit, the normalized root-mean-squared errors (NRMSEs) were calculated

in both cases. Repeated experiments were performed to investigate if the models for fitting data obtained during the first day of collecting signals could be successfully used on the data from the remaining days.

2. Materials and Methods

2.1. Subjects

Five healthy non-smoker volunteers (four males and one female), with the mean age and standard deviation of 27 ± 7.5 years, weight of 63.5 ± 5 kg, and height 173.2 ± 8.4 cm, were asked to participate in this study. Individuals with previous pneumothorax, chronic respiratory illnesses, and common cold were excluded from the study. This group of participants consisted of students and staff members from the University of Connecticut (UConn, Storrs, CT, USA). All participants signed a consent form approved by the Institutional Review Board of UConn.

2.2. Equipment and Acquisition of the Signals

In this study, two signals were acquired simultaneously: tracheal sounds and Resptrace signal. The tracheal sounds were collected using an acoustical sensor, which contained a subminiature electret microphone BT-21759-000 (Knowles Electronics, Itasca, IL, USA) placed in a plastic bell, which consisted of a conical coupler chamber [25], in accordance to previous findings [26]. The importance of this shape is that it provides an efficient transducer of air pressure fluctuations from the skin over the trachea to the microphone [27]. The acoustic sensor used in this study was developed by our colleagues at the Metropolitan Autonomous University at Mexico City, Mexico, and have been successfully applied for respiratory sound acquisitions [18,25,28]. The acoustic sensor was connected to the audio jack of the Samsung Galaxy S4 smartphone (Samsung Electronics Co., Seoul, Korea). The tracheal sounds were recorded using the built-in audio recorder application (Voice Recorder), with 16-bit per sample and 44.1 kHz sampling rate, and saved in the .wav format. Afterwards, the recorded files were transferred to a personal computer and processed offline using Matlab (R2012a, The Mathworks, Inc., Natick, MA, USA).

The Resptrace (nowadays known as Inductotrace) signal was obtained simultaneously with the tracheal sounds, from two Respibands (Ambulatory Monitoring, Inc., Ardsley, NY, USA), placed over the rib cage and abdomen. Respibands' signals were digitized using 16-bit A/D converter (PowerLab/4SP, ADInstruments, Inc., Dunedin, New Zealand) at 10 kHz sampling rate, using the manufacturer's software (LabChart 7, ADInstruments, Inc.). Prior to every participant's recording, the Respibands were calibrated using a spirometer system (FE141 Spirometer, ADInstruments, Inc.) following the manufacturer's manual, and the corresponding signal was considered as the reference for volume estimation. Calibration errors between Respibands and spirometer were obtained for every recording, and were less than 10%, which is in accordance to the manufacturer's manual.

Experiments were performed in a regular dry lab which was held quiet. Respibands were placed over the participant's rib cage and abdomen, while the acoustical sensor was fixed at the suprasternal notch using a double-sided adhesive ring (BIOPAC Systems, Goleta, CA, USA). The experiment consisted of three stages, and all were performed in standing posture:

1. Participants were asked to breathe through an 800 mL Spirobag (Ambulatory Monitoring, Inc., Ardsley, NY, USA) for about six respiratory cycles;

2. Participants were asked to follow a maneuver that consisted of increasing tidal volumes and then decreasing with each breath, ranging from participant's comfortable lowest to highest volume, while breathing through a paper tube (tube's length: 20 cm, internal diameter: 1.5 cm, external diameter: 2 cm), for approximately 2 min;
3. Participants were asked to repeat the same maneuver as in the second stage while breathing without the tube.

In everyday situations people do not have access to spirometers or Respibands, and the lack of portable and easily accessible device with possibility to control and limit the tidal volume is needed. Thus, in this research, we use a Spirobag, since it is easy to find and carry, and has an almost fixed volume (800 mL). The exact volume of the bag changes at each volunteers' breathe. Hence, we used the Resptrace system as reference in order to know this volume, since the use of spirometer with a bag was practically prohibited in the experimental setup.

Since breathing through a tube adds some resistance to the respiratory tract and changes the natural way of breathing, one of the objectives was to investigate if this apparatus influences the estimation results. This was the reason for recording the third stage of the experiment.

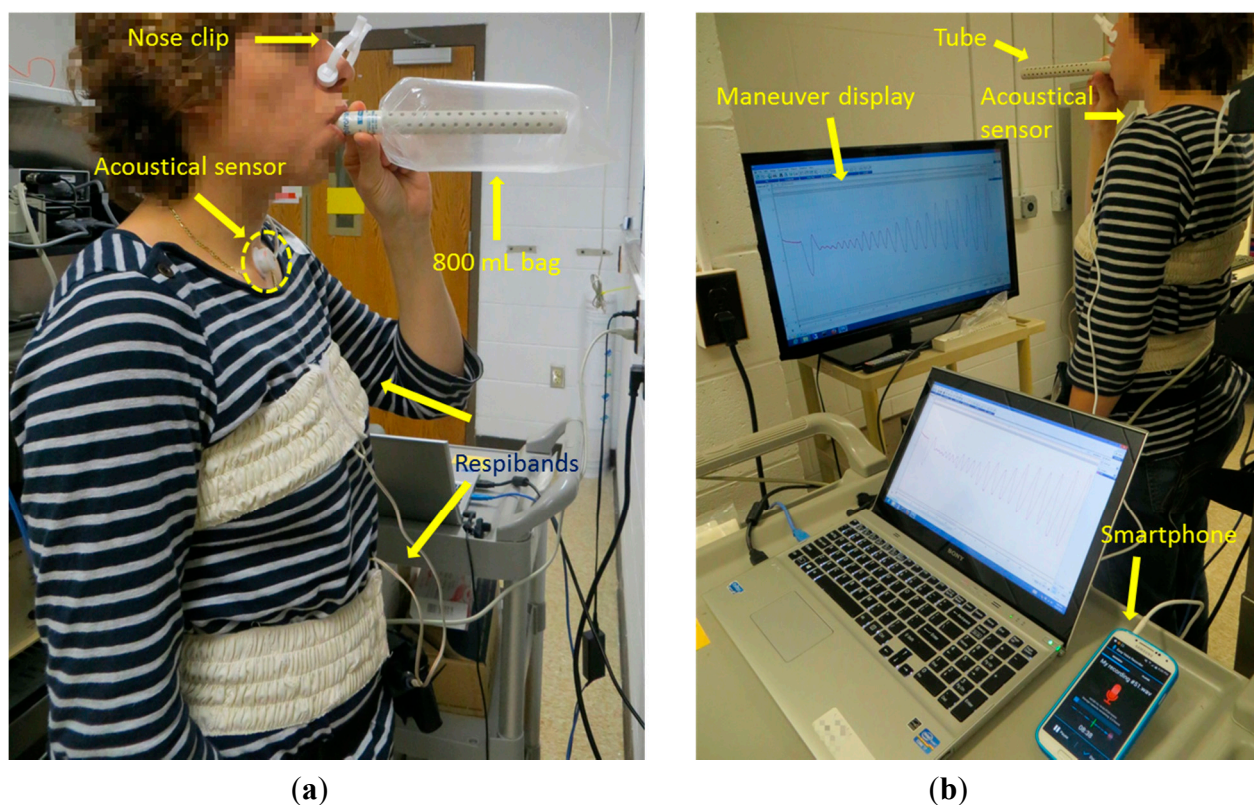


Figure 1. Simultaneous recordings of the tracheal sound (using a smartphone) and the volume signal (using Respibands). (a) The participant is breathing through 800 mL bag; (b) The participant is breathing through a tube while performing the respiratory maneuver.

In all three stages, initial and final apnea phases of approximately 5 s were acquired for automatic alignment purposes between the two recordings, as well as for recording the ambient noise levels. In the last two stages, after the initial apnea, participants were instructed to take a forced respiration cycle before performing the maneuver. In order to provide the visual feedback during the second and the third

stage, the volume signal was displayed on a 40" monitor, placed in front of the participant. During the experiment, nose clips (MLA1008, ADInstruments, Inc.) were used to clamp the nostrils. An example of the set-up of the experiment is shown in Figure 1. Figure 1a depicts the first stage of the experiment, when the 800 mL bag was used, while Figure 1b shows the breathing maneuver through a tube (the second stage of the experiment).

2.3. Data Processing

Figure 2 shows the flowchart of the data processing steps. The acquired tracheal sounds were first downsampled from 44.1 kHz to 6.3 kHz, and then digitally filtered with a 4th order bandpass Butterworth filter with cutoff frequencies 100 and 3000 Hz to minimize the effects of heart sounds and muscle interferences [27,29]. The volume signal was first downsampled from 10 kHz to 5 kHz, and then interpolated to 6.3 kHz in order to achieve the same sampling frequency as the tracheal sounds. Lastly, the volume signal was lowpass filtered at 2 Hz with a 4th order Butterworth filter.

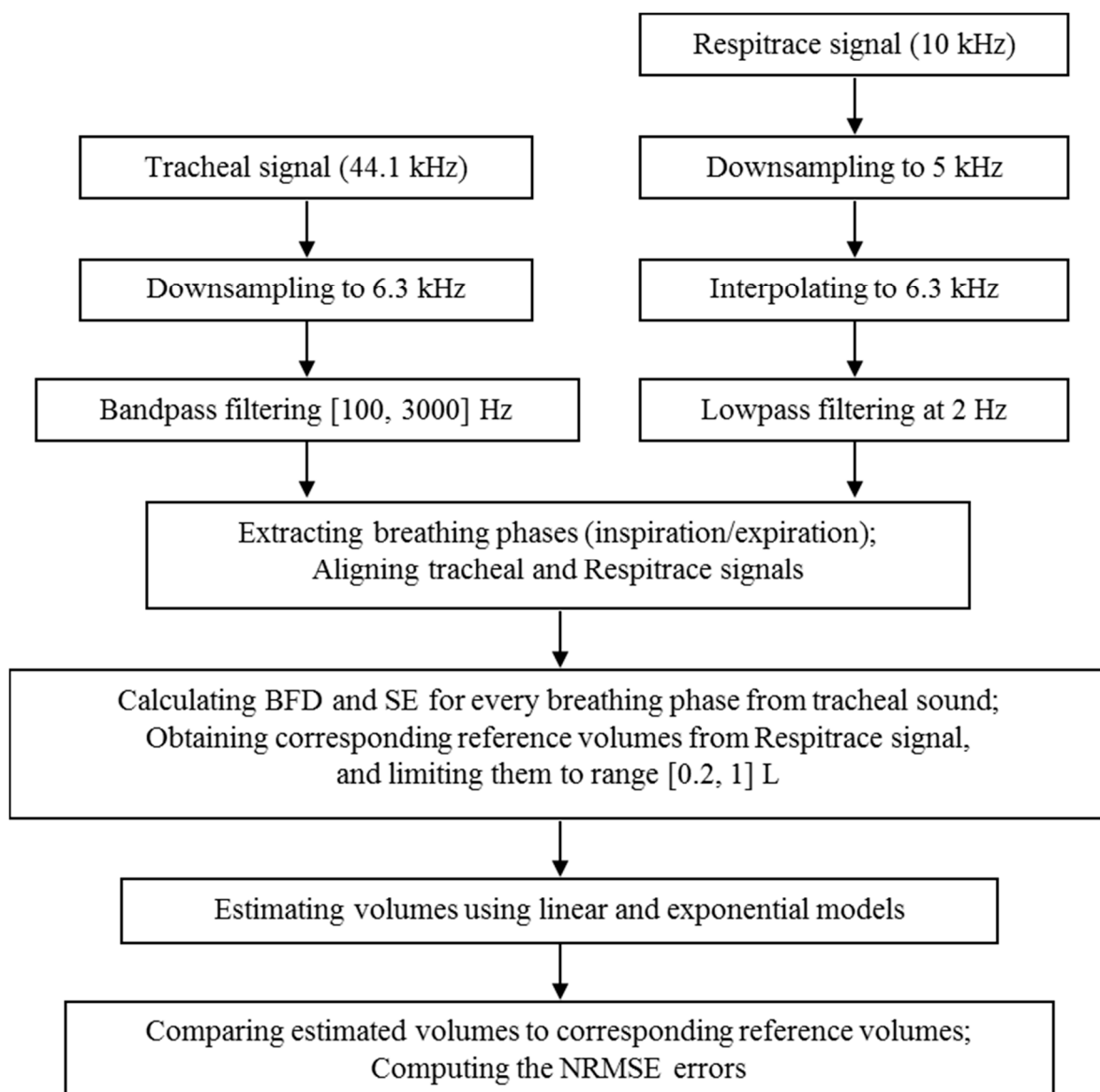


Figure 2. The flowchart showing the steps for tracheal sounds' and Respiration signal's processing.

The automatic extraction of the breathing phases (inspiration/expiration) was performed from the volume signal, by finding its corresponding local maxima and minima during the respiratory maneuver and computing the slope of the volume at each phase [18]. The tracheal sounds and the volume signal were recorded simultaneously, however, due to the different times of pressing the start buttons, the two signals were aligned manually. Figure 3 depicts an example of the filtered, detrended and aligned tracheal sounds and volume signal during the respiratory maneuver.

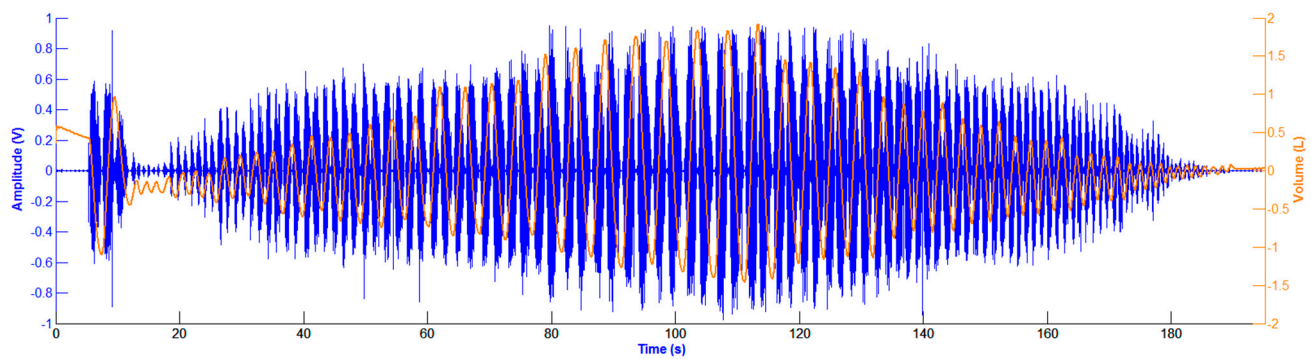


Figure 3. Filtered, detrended and aligned tracheal sounds and volume signal during the respiratory maneuver. Tracheal sound (in volts) is represented in blue, while volume signal (in liters) is in orange.

The volume signal, acquired with the Respibands, was assumed as the reference. For every breathing phase, the absolute volume difference between two consecutive extrema from the volume signal was calculated, and was considered as the true tidal volume value, V_T . Two features were used for estimating the tidal volume from the tracheal sounds acquired by smartphone: blanket fractal dimension (BFD) and the integral of the Shannon entropy (SE). Every breathing phase (inspiration/expiration) from the tracheal sound was represented with one BFD and one SE value. In order to estimate the volume from these features, linear and exponential fitting curves were used. The estimated volumes are defined with the following:

$$\begin{aligned} V_{est_l} &= a \cdot F + b \\ V_{est_e} &= c \cdot e^{d \cdot F} \end{aligned} \quad (1)$$

where V_{est_l} and V_{est_e} are the estimated volumes with linear and exponential models, respectively, a , b , c and d are coefficients, and F is the value of the BFD or SE feature computed from the tracheal sounds.

The last step in the data processing is the comparison of the estimated volumes to the corresponding reference volume values, and the evaluation of the performed estimation via computation of the normalized root-mean-squared error (NRMSE) defined as follows:

$$\begin{aligned} NRMSE &= \frac{RMSE}{\text{mean}(V_T)} \cdot 100\% \\ RMSE &= \sqrt{\frac{\sum_{i=1}^N (V_T(i) - V_{est}(i))^2}{P}} \end{aligned} \quad (2)$$

where V_T is the volume obtained from Resptrace, V_{est} denotes the estimated volume, *i.e.*, V_{est_l} or V_{est_e} , and P is the number of breathing phases during the maneuver.

Shannon entropy is a measure of uncertainty or irregularity of a process [30]. It is one of the features frequently used for analysis of respiratory sounds, and has been successfully applied to airflow estimation in the field of tracheal sound analysis [31]. For a random signal with a probability density function (pdf), p , SE is defined as:

$$SE(p) = -\sum_{i=1}^M p_i \cdot \log p_i \quad (3)$$

where M is the number of outcomes of the random variable with pdf p . In this study, pdf is estimated using the method of Parzen's windows with a Gaussian kernel [32,33]. More details on this method can be found in [18,31]. In this study we were concerned with the tidal volume estimation rather than respiratory airflow, and based on the relationship between these two variables over time, the integral of the SE over each corresponding breathing phase was used as feature for tidal volume estimation.

2.4. Blanket Fractal Dimension

Fractals are defined as 'a set having the fractal dimension strictly greater than its integer dimension', and are used to describe non-regular and non-stationary structures [34–36]. There are two types of fractals: natural and deterministic. Natural fractals are structures that could be found in the nature, such as lungs, while deterministic fractals are constructed artificially, by applying predetermined replicating rules (e.g., the Von Koch curve, the Cantor set) [36,37]. Fractal structures may be quantified by fractal dimension, which is a number (usually non-integer) expressing the manner in which the irregular structure replicates itself through different scales [36,37]. Among various fractal dimensions, in this study we used blanket fractal dimension (BFD). The BFD was initially proposed for estimating fractal dimension of digital images (2D signals) [38], and is further extended to 1D signals [39].

In the case of 1D signals, the set of points within maximal distance ε from a curve is considered. Therefore, a strip of width 2ε that surrounds the curve is observed [40]. Blanket method creates the strip around the signal, defined by the upper and lower limiting lines, defined as follows [39]:

$$\begin{aligned} u_{\varepsilon}(i) &= \max \left\{ u_{\varepsilon-1}(i) + 1, \max_{|m-i| \leq 1} u_{\varepsilon-1}(m) \right\} \\ b_{\varepsilon}(i) &= \min \left\{ b_{\varepsilon-1}(i) - 1, \min_{|m-i| \leq 1} b_{\varepsilon-1}(m) \right\} \\ u_0(i) &= b_0(i) = x(i) \end{aligned} \quad (4)$$

where $x(i)$ represents the observed 1D signal, $u_{\varepsilon}(i)$ and $b_{\varepsilon}(i)$ are the upper and lower lines, respectively, i is the current sample of the signal, m denotes samples within the window around the current sample of the signal, and ε is the predefined maximal distance of upper/lower line from the signal. As can be noted from Equation (4), the upper/lower line is always calculated for the three consecutive samples: $i-1$, i , and $i+1$.

The area of the strip between upper and lower lines is defined as:

$$A_{\varepsilon} = \sum_i \{u_{\varepsilon}(i) - b_{\varepsilon}(i)\} \quad (5)$$

from which the length of the curve x can be estimated as [39]:

$$L(\varepsilon) = \frac{A_\varepsilon - A_{\varepsilon-1}}{2} \quad (6)$$

On the other hand, the length of the curve follows the power law [36]:

$$L(\varepsilon) = C \cdot \varepsilon^{1-D} \quad (7)$$

where C is the constant and D is the blanket fractal dimension (BFD). By combining Equations (6) and (7), and using the least square approximation, blanket fractal dimension is calculated.

3. Results

All five participants performed the experiments described in Section 2.2 six times in six distinct days, thus creating a database of 30 recordings. The data collected on the first day were used for obtaining the linear and exponential models, while the data from the remaining five days were used for testing the previously obtained models. Each breathing phase, inspiration and expiration, was analyzed separately.

The linear and exponential fitting curves were calculated only from the first stage of the experiment performed during the first day, using two and three points, respectively, when the participant was breathing through an 800 mL bag for about six respiratory cycles. BFD and SE features were calculated from the smartphone acquired tracheal sounds, while the reference volume values were obtained from the Resptrace signal. This was performed for every inspiratory and expiratory phase, as well as for the portion of the signal during the initial apnea (denoted as background). For the linear fitting curve, for both BFD and SE features, it was found, experimentally, that two points, A and B , with the following coordinates:

$$\begin{aligned} A &= (x_1, y_1) = (\text{mean}(\text{feature values for 800 mL}), \text{mean}(\text{volumes for 800 mL})) \\ B &= (x_2, y_2) = (\text{feature value of background for 800 mL}, \text{volume of background for 800 mL}) \end{aligned} \quad (8)$$

are sufficient for determining the fitting line.

Similarly, for exponential fitting curves, we found empirically that three points are sufficient, as follows. When using BFD features, the three points (C, D, E) are:

$$\begin{aligned} C &= (x_3, y_3) = (\text{mean}(\text{BFD for 800 mL}), \text{mean}(\text{volumes for 800 mL})) \\ D &= (x_4, y_4) = (0.8, 0) \\ E &= (x_5, y_5) = (2, 2) \end{aligned} \quad (9)$$

and with SE features (points F, G, H):

$$\begin{aligned} F &= (x_6, y_6) = (\text{mean}(\text{SE for 800 mL}), \text{mean}(\text{volumes for 800 mL})) \\ G &= (x_7, y_7) = (0, 0.2) \\ H &= (x_8, y_8) = (6, 2) \end{aligned} \quad (10)$$

After investigating values of the BFD and SE features from all participants, we noticed that the upper limits were 2 and 6, for BFD and SE respectively. Therefore, we used these asymptotic values as abscissae of points E and H . Figure 4 illustrates the computation of the linear and exponential models.

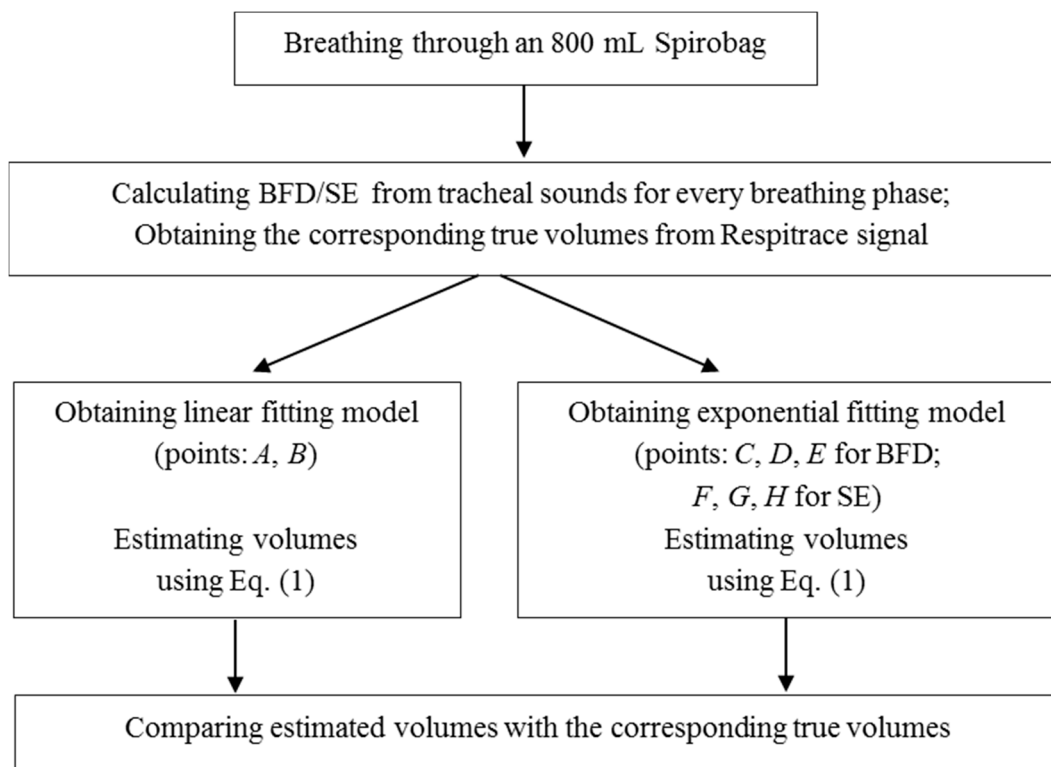


Figure 4. The flowchart showing the computation of the fitting models.

After the linear and exponential curves are calculated, data from the second and the third stages of the experiment (breathing with and without a tube) were used to fit the curves, separately. BFD and SE features were calculated from the smartphone acquired tracheal sounds, and the corresponding volumes were estimated using Equation (1) for the linear and exponential models. Simultaneously, the true volume values were obtained from the reference Resptrace signal. Since the volume range for normal breathing is between 0.2 and 1 L [7], we limited the true volume values to this range, and used only the corresponding portions of tracheal sounds for analysis.

An example of the volume estimation from smartphone acquired tracheal sounds using BFD features and exponential model, for both inspiration and expiration, of one subject is shown in Figure 5. The true tidal volume values (from Resptrace system) and their corresponding BFD values when breathing through 800 mL bag and tube are represented in blue squares and green circles, respectively, while the estimated volumes and their corresponding BFD features are depicted as brown triangles. The three points, shown as black marks in Figure 5 and given with Equation (9), are used for obtaining the exponential fitting curve, which is shown as a solid red curve.

For every inspiration and expiration phase, when a true volume value was between 0.2 and 1 L, the estimated volumes were compared to their corresponding true volumes, and NRMSEs were calculated using Equation (2). In Figure 6 are shown the estimated and reference volumes, as well as the corresponding NRMSE errors for every inspiratory and expiratory phase for the same example as in Figure 5.

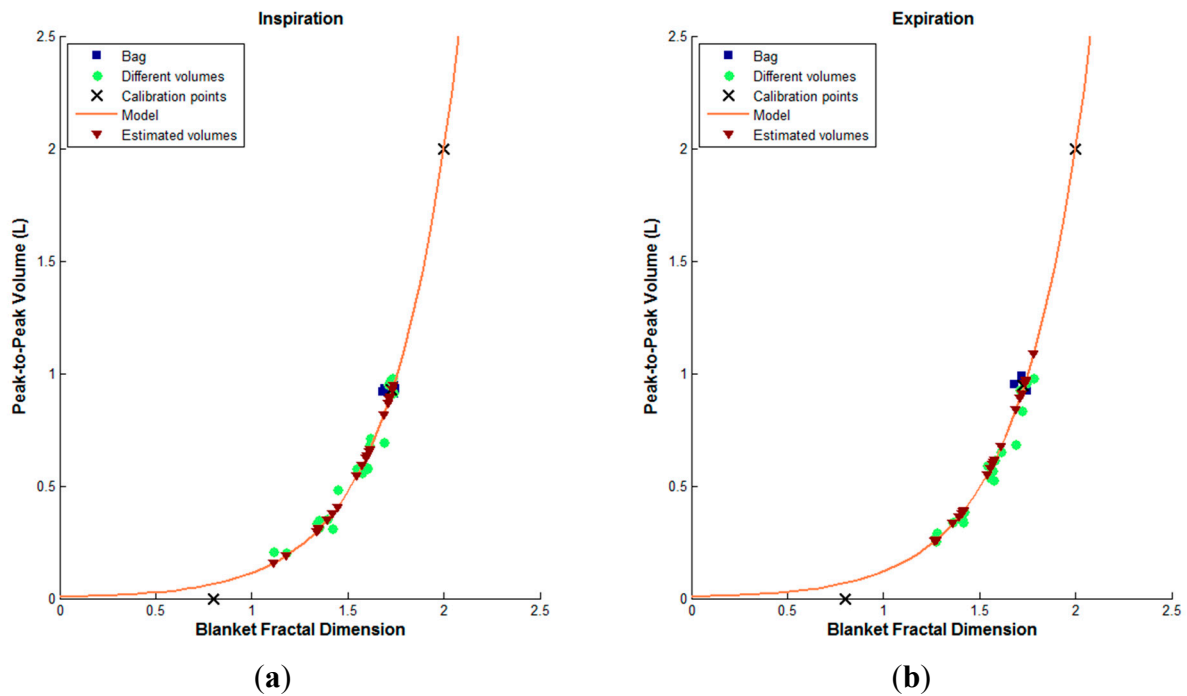


Figure 5. An example of the volume estimation from smartphone acquired tracheal sounds using BFD features and exponential model of one subject. The true volumes while breathing through a tube (green circles) are limited to a range from 0.2 to 1 L. (a) The inspiration phase; (b) The expiration phase.

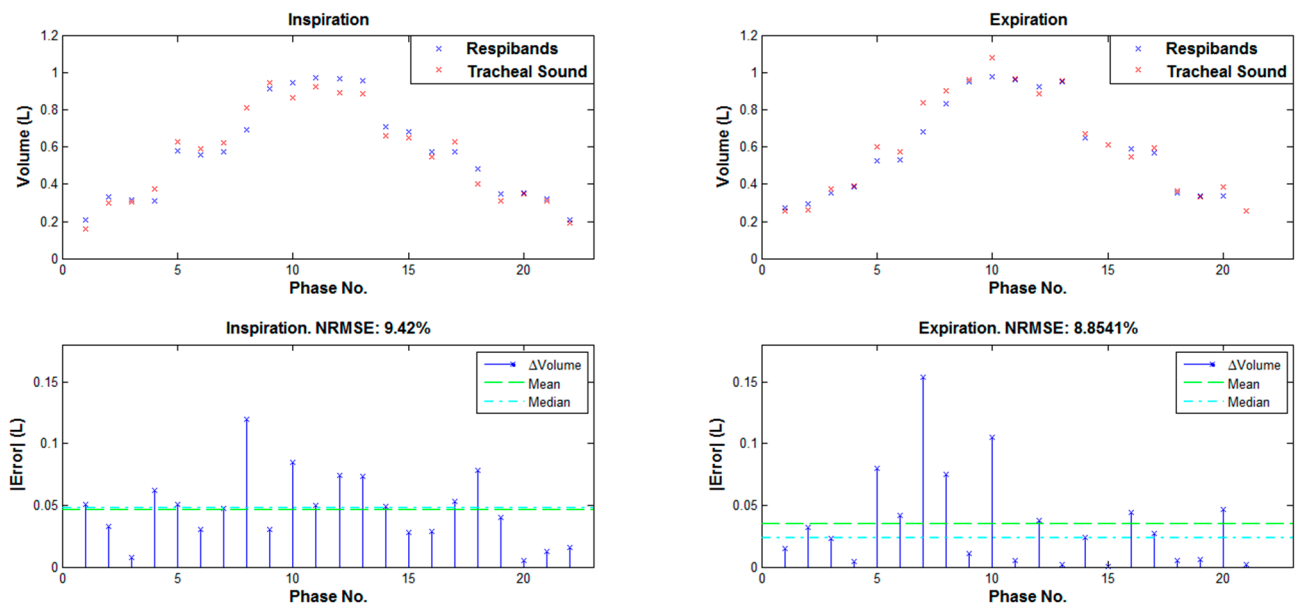


Figure 6. Top: Reference and estimated volumes for the same example as in Figure 5. **Bottom:** The corresponding NRMSE errors.

As can be noted from Figure 6, values of the volumes estimated from a smartphone acquired tracheal sounds using the BFD features are very similar to the volume values obtained from a RespiTrace (reference) signal; and the NRMSE errors in both inspiration and expiration phases are low (less than 10%).

After the first day of experiments (later denoted as training), the participants repeated breathing maneuvers with and without a tube for five days (denoted as tests 1–5). The BFD and SE features were calculated from the tracheal sounds, and the volumes were estimated using the first day's fitting curves. Simultaneously, the true volume values were obtained from the Resptrace signal. Again, the estimated volumes were compared to the true volumes, and NRMSEs were calculated.

In this study, we compared the volume estimation results when the proposed blanket fractal dimension is used as feature, with results obtained with Shannon entropy. Conditions of comparisons included: the type of the model (exponential, linear), the type of the apparatus (tube, no tube), and the breathing phase (inspiration, expiration). All combinations of conditions were made, and the corresponding ones were tested statistically, using the two-tailed paired *t*-tests (SPSS Statistics 20, IBM Corporation, Armonk, NY, USA). Table 1 contains the list of combinations and their corresponding *p*-values when statistically significant differences occurred ($p < 0.05$).

Table 1. Combinations of conditions when statistically significant differences were obtained, and their corresponding *p*-values. Results are grouped into 4 groups, based on the type of comparisons performed, *i.e.*, BFD vs. SE; inspiration vs. expiration; no tube vs. tube; exponential vs. linear model.

Type	Day	Conditions	<i>p</i> -value
BFD vs. SE	Test 4	Exponential, tube, inspiration	0.049
		Exponential, tube, expiration	0.015
		Exponential, no tube, expiration	0.011
		Linear, tube, inspiration	0.037
		Linear, tube, expiration	0.013
		Linear, no tube, expiration	0.002
	Test 5	Exponential, tube, expiration	0.017
		Linear, tube, expiration	0.006
		Linear, no tube, expiration	0.007
Inspiration vs. Expiration	Test 1	BFD, linear, tube	0.033
	Test 4	SE, linear, tube	0.025
	Test 5	BFD, linear, tube	0.022
		SE, exponential, tube	0.029
		SE, linear, tube	0.031
No tube vs. Tube	Training	SE, exponential, inspiration	0.016
	Test 4	BFD, linear, inspiration	0.042
	Test 5	BFD, linear, inspiration	0.033
Exponential vs. Linear	Training	BFD, tube, expiration	0.008
		BFD, no tube, expiration	0.038
	Test 4	BFD, tube, expiration	0.028
		SE, tube, expiration	0.018
	Test 5	SE, tube, expiration	0.028

In addition, for each combination, the comparisons between results (NRMSE errors) of the training day and the five test days were performed, and tested statistically using the repeated measures ANOVA with Bonferroni *post-hoc* tests (SPSS Statistics 20). The NRMSE errors are grouped into four parts, based on the apparatus and breathing phase, so that comparisons between features and models can be

performed, and are depicted in Figure 7. These graphs show the changes in NRMSE errors throughout six days of experiments for all combinations of features and models simultaneously.

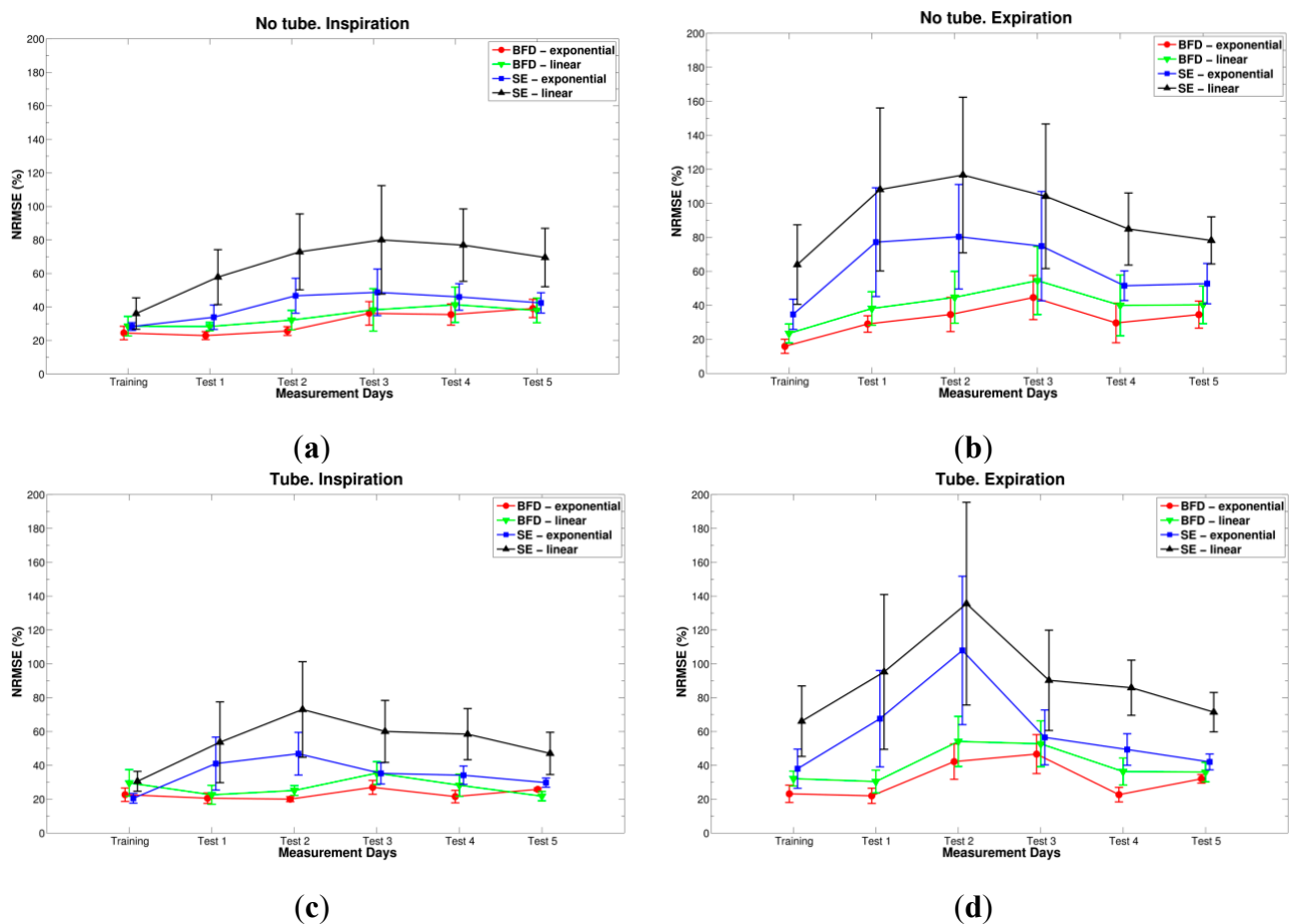


Figure 7. NRMSE errors (represented with its mean and standard error of the mean) when: BFD and exponential model (red circles), BFD and linear model (green downward triangles), SE and exponential model (blue squares), and SE and linear model (black triangles) are used. (a) No tube and inspiration; (b) No tube and expiration; (c) Tube and inspiration; (d) Tube and expiration.

As can be concluded from the graphs in Figure 7, when blanket fractal dimension was used for volume estimation (red and green lines), the errors were lower at least two times than when Shannon entropy was used (blue and black lines), especially with the exponential model (red circles). Moreover, note that standard errors are also smaller when BFD is used. Statistically significant differences between the two features appeared during the fourth test day (for: exponential and linear models, with tube and both inspiration and expiration phases; and for both models, without tube and expiration) and the fifth test day (for: both models, with tube and expiration phase; and linear model, without a tube and expiration), as shown in Table 1.

The smallest NRMSE error, with mean and standard deviation of $15.877\% \pm 9.246\%$, was obtained during the first day of experiments (training), when BFD feature with the exponential model was used, for expiratory phase, while the participants were breathing without a tube, Figure 7b. The Bland-Altman

analysis showed a bias and standard deviation of 0.0226 ± 0.0918 L, and the corresponding results are presented in Figure 8.

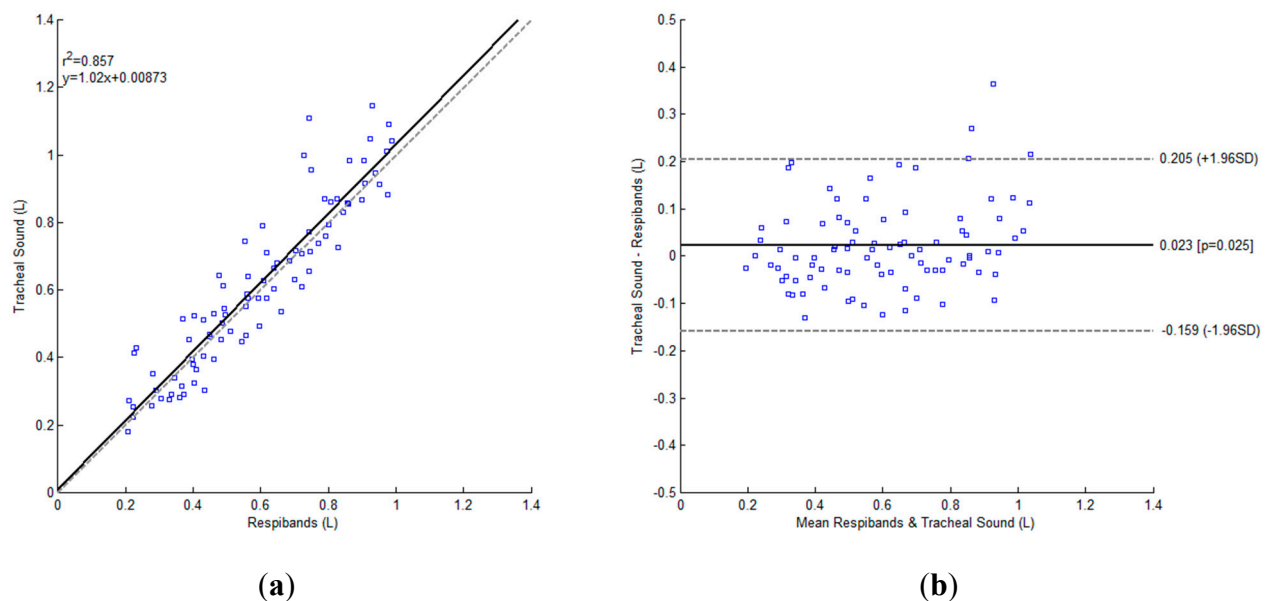


Figure 8. Bland-Altman plot for BFD feature with the exponential model, for expiratory phase, while the participants ($N = 5$) were breathing without a tube during the first day of experiments. **(a)** The regression plot: The unitary line is shown as gray dashed line, while the regression line is represented as black solid line; **(b)** Bland-Altman plot: The bias is represented as a solid black line and the 95% limits of agreement as gray dashed lines.

By looking at the NRMSEs calculated for the remaining 5 days (test days), one can conclude that the smallest was always obtained with the BFD feature, exponential model and inspiration while breathing through a tube (errors ranging from 20% to 27%), Figure 7c, except for the fifth day, when linear model provided better estimation (error around 21%). No statistically significant differences were found between BFD exponential model from inspiratory and expiratory phases, as deduced from Table 1.

As was mentioned above, when BFD feature was used the errors were always smaller than with SE. In addition, one can conclude that the fitting curves obtained during the first day of experiments (training) can be successfully used for the following test days. This way, the participants do not need to perform all three stages of the experiments, and the fitting curves do not need to be calculated every day, as the previously determined could be used. In order to statistically compare errors throughout all six days of experiments, repeated measures ANOVA with Bonferroni *post-hoc* tests were performed, and was determined that there were no statistically significant differences between the days of experiments when BFD or SE was used as feature. According to Table 1, for the BFD using exponential model, no statistically significant differences were found between breathing through the tube or not.

4. Discussions and Conclusions

The goal of this study was to estimate tidal volume from the smartphone acquired tracheal sounds. The main challenge was to find a suitable feature to describe these sounds, such that the volume

could be estimated directly from the sounds as accurate as possible. Respiratory sounds, and hence tracheal sounds, are non-stationary and stochastic signals [2], and as such they are suitable for fractal analysis [36]. We tested several ways for estimating fractal dimension, and decided to use the blanket fractal dimension because it was more suitable for describing and following the dynamics of the tracheal sounds, which was evident after exploring the results. Possible explanation could be the definition of the blanket fractal dimension itself. Blanket method creates a strip around the tracheal signal, closely following the changes in the signal. As the signal changes faster, the value of blanket fractal dimension becomes higher. In some past studies fractal analysis and fractal dimensions were used for analyzing tracheal and lung sounds [20–24]. Moreover, blanket fractal dimension was not used in respiratory sound analysis yet, and especially not for estimating the tidal volume, which are some of the novelties of this manuscript. In addition, to the best of our knowledge, none of the studies on tidal volume estimation has reported results based on tracheal sounds acquired by a smartphone.

In addition to BFD features, we used Shannon entropy (SE), as it is one of the features frequently used for analysis of respiratory sounds. In [41], the authors proposed a method to estimate airflow from tracheal sounds using SE. In [42], the authors proposed tidal volume estimation method by integrating airflow derived from tracheal sounds, which takes advantage of airflow/sound intensity relationship. As can be noted, the straightforward comparison between our method and method used in [42] is difficult to perform, since the conditions are not exactly the same. We estimated the tidal volume directly from tracheal sounds, using BFD as a feature, while Que *et al.* [42] obtained first the relationship between sounds' amplitude and airflow, and then the volume by integrating the flow. Consequently, according to the provided results, the range of volume values in [42] was roughly between 0.3 and 0.8 L, while we limited volumes to a broader range [0.2, 1] L. That being said, the Bland-Altman analysis results of [42] were 0.009 ± 0.046 L (bias $\pm$ SD), while we found a bias and standard deviation of 0.0226 ± 0.0918 L. Chen *et al.* estimated tidal volume from the energy of the tracheal sounds [6]. The comparison of our results with those reported in [6] is not easy to perform since they are reported separately for each individual participant. If we compute the average results from the provided individually-based values reported in Table 1 [6]), we can conclude that the results are comparable. The volumes ranged from 0.15 to 0.5 L in [6], which is notably smaller range than the one used in this study. Note that in contrast to these two studies, the only external information needed to compute the calibration model with our proposed method was obtained with a simple bag at a known fixed value and not from a spirometer-like device.

After volumes were estimated from the smartphone acquired tracheal sounds, they were compared to the true volume values, obtained from RespiTrace signal, which was considered as a reference in this study. The RespiTrace signal was calibrated against the spirometer signal prior every recording and the obtained calibration errors were less than 10%, which is in accordance to the manufacturer's manual. These reference volumes were limited to a range from 0.2 to 1 L, as it is the normal breathing range [7]. Inspiratory and expiratory phases were analyzed separately. Two fitting models, exponential and linear, were used for estimation. Our results indicate that the best estimation was obtained using blanket fractal dimension with exponential model, during expiratory phase, while participants were breathing without a tube, when the NRMSE error was $15.877\% \pm 9.246\%$ (expressed as mean $\pm$ standard deviation). In addition, when the BFD is used as a feature, the NRMSEs were always smaller, at least twice, compared to the SE.

The experiments involved acquisition during six days. Data from the first day of experiments were used to construct estimation models, while the data from the remaining five days were plotted against the obtained models. The results show the possibility to successfully apply previously obtained fitting curves and to monitor tidal volume for at least five days. This way we introduce an easy calibration procedure, where there is no need to calculate fitting curves prior every consecutive experiment. In our future work, we plan to determine for how many days the existing models can be used.

This is a preliminary study, with the objective to estimate tidal volume in healthy participants, and not in patients with pulmonary diseases. Therefore, it was performed on five healthy participants, and for the future work we plan to expand the group. This study was limited to acquisition of tracheal sounds in standing posture without head movements. We expect that the results obtained with the proposed methodology would be in agreement with the study reported in [42], where the effects of body movements and posture changes on tidal volume estimates were investigated. Accordingly, we foresee that head movements without neck extension will not modify the obtained results and we do not anticipate an increase in estimation errors when moving to seated posture, but we do when moving from standing to supine posture, where a new calibration in latter posture would be required. It is worth to mention that all recordings were made in a regular dry lab, that was held quiet, and not in a special soundproof environment, hence making it applicable to real-life situations. Since spirometer is not a portable device, not easily accessed and fixed values of tidal volumes are hard to control, which results in additional turbulences and changes in breathing patterns, we used a Spirobag in order to obtain information at a known volume which in turn was employed in the estimation model. In addition, due to high performance capabilities of smartphones, by connecting an adequate acoustical sensor to a smartphone and using a Spirobag, a portable system for tidal volume estimation can be obtained.

In summary, in this manuscript we proposed a novel technique for estimation of tidal volume directly from the blanket fractal dimension of the tracheal sounds. The proposed method provided promising results and outperformed a method based on the Shannon entropy, which is frequently used in tracheal sounds analysis. Furthermore, we introduced an easy calibration procedure that does not require specialized devices and when combined with the proposed signal processing technique allows reasonable estimation for at least five days, which makes this method easier to use in everyday situations. The employment of smartphone-acquired tracheal sounds was also introduced for all of the above mentioned purposes. We foresee that similar efforts to the one presented here represent a step forward to the development of a mobile breathing monitoring system easily available for the general population.

Acknowledgments

This work is supported in part by the US Army Medical Research and Materiel Command (US-AMRMC) under grant No. W81XWH-12-1-0541. The authors would like to thank Sonia Charleston-Villalobos, Thomas Aljama-Corrales, and Ramon Gonzales-Camarena for introducing a microphone sensor used in this research work. The authors would also like to thank Milorad Paskas for his help in programs for calculating blanked fractal dimension. We also acknowledge the feedback provided by the reviewers and editor to improve the quality of the paper.

Author Contributions

Natasa Reljin and Bersain Reyes designed and performed the experiments, analyzed the data, and wrote the manuscript. Ki Chon is the Principal Investigator, and oversaw and guided the study and the writing of the manuscript.

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. Moussavi, Z. *Fundamentals of Respiratory Sounds and Analysis*, 1st ed.; Morgan & Claypool Publishers: San Rafael, CA, USA, 2006.
3. Sovijarvi, A.R.A.; Malmberg, L.P.; Charbonneau, G.; Vanderschoot, J.; Dalmasso, F.; Sacco, C.; Rossi, M.; Earis, J.E. Characteristics of breath sounds and adventitious respiratory sounds. *Eur. Respir. Rev.* **2000**, *10*, 591–596.
4. Folke, M.; Cernerud, L.; Ekström, M.; Hök, B. Critical review of non-invasive respiratory monitoring in medical care. *Med. Biol. Eng. Comput.* **2003**, *41*, 377–383.
5. Kuratomi, Y.; Okazaki, N.; Ishihara, T.; Arai, T.; Kira, S. Variability of breath-by-breath tidal volume and its characteristics in normal and diseased subjects. *Jpn. J. Med.* **1985**, *24*, 141–149.
6. Chen, G.; de la Cruz, I.; Rodriguez-Villegas, E. Automatic lung tidal volumes estimation from tracheal sounds. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, IL, USA, 26–30 August 2014; pp. 1497–1500.
7. Sherwood, L. *Fundamentals of Human Physiology*, 4th ed.; Cengage Learning: Boston, MA, USA, 2012.
8. Semmes, B.J.; Tobin, M.J.; Snyder, J.V.; Grenvik, A. Subjective and objective measurement of tidal volume in critically ill patients. *Chest* **1985**, *87*, 577–579.
9. Grossman, P.; Spoerle, M.; Wilhelm, F.H. Reliability of respiratory tidal volume estimation by means of ambulatory inductive plethysmography. *Biomed. Sci. Instrum.* **2006**, *42*, 193–198.
10. Sayadi, O.; Weiss, E.H.; Merchant, F.M.; Puppala, D.; Armoundas, A.A. An optimized method for estimating the tidal volume from intracardiac or body surface electrocardiographic signals: Implications for estimating minute ventilation. *Am. J. Physiol. Heart Circ. Physiol.* **2014**, *307*, H426–H436.
11. Corbishley, P.; Rodriguez-Villegas, E. Breathing detection: Towards a miniaturized, wearable, battery-operated monitoring system. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 196–204.
12. Earis, J.E.; Cheetham, B.M.G. Future perspectives for respiratory sound research. *Eur. Respir. Rev.* **2000**, *10*, 641–646.

13. Cala, S.J.; Kenyon, C.M.; Ferrigno, G.; Carnevali, P.; Aliverti, A.; Pedotti, A.; Macklem, P.T.; Rochester, D.F. Chest wall and lung volume estimation by optical reflectance motion analysis. *J. Appl. Physiol.* **1996**, *81*, 2680–2689.
14. Petrovic, M.D.; Petrovic, J.; Danicic, A.; Vukcevic, M.; Bojovic, B.; Hadzievski, Lj.; Allsop, T.; Lloyd, G.; Webb, D.J. Non-invasive respiratory monitoring using long-period fiber grating sensors. *Biomed. Opt. Express* **2014**, *5*, 1136–1144.
15. Lee, Y.S.; Pathirana, P.N.; Steinfort, C.L.; Caelli, T. Monitoring and analysis of respiratory patterns using microwave doppler radar. *IEEE J. Transl. Eng. Health Med.* **2014**, *2*, 1–12.
16. 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.
17. 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.
18. Reyes, B.A.; Reljin, N.; Chon, K.H. Tracheal sounds acquisition using smartphones. *Sensors* **2014**, *14*, 13830–13850.
19. Ahlstrom, C.; Johansson, A.; Hult, P.; Ask, P. Chaotic dynamics of respiratory sounds. *Chaos Solitons Fractals* **2006**, *29*, 1054–1062.
20. Yap, Y.L.; Moussavi, Z. Respiratory onset detection using variance fractal dimension. In Proceedings of 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey, 25–28 October 2001; pp. 1554–1556.
21. Gnitecki, J.; Moussavi, Z. Variance fractal dimension trajectory as a tool for heart sound localization in lung sounds recordings. In Proceedings of 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, 17–21 September 2003; pp. 2420–2423.
22. Hadjileontiadis, L.J.; Rekanos, I.T. Detection of explosive lung and bowel sounds by means of fractal dimension. *IEEE Signal Process. Lett.* **2003**, *10*, 311–314.
23. Gnitecki, J.; Moussavi, Z. The fractality of lung sounds: A comparison of three waveform fractal dimension algorithms. *Chaos Solitons Fractals* **2005**, *26*, 1065–1072.
24. Hadjileontiadis, L.J. A novel technique for denoising explosive lung sounds: Empirical mode decomposition and fractal dimension filter. *IEEE Eng. Med. Biol. Mag.* **2007**, *26*, 30–39.
25. 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.* **2013**, *52*, 266–276.
26. Druzgalski, C.K.; Donnerberg, R.L.; Campbell, R.M. Techniques of recording respiratory sounds. *J. Clin. Eng.* **1980**, *5*, 321–330.
27. Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory sounds: Advances beyond the stethoscope. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987.
28. 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.

29. Reichert, S.; Gass, R.; Brandt, C.; Andres, E. Analysis of respiratory sounds: State of the art. *Clin. Med. Insights Circ. Respir. Pulm. Med.* **2008**, *2*, 45–58.
30. Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill: New York, NY, USA, 2002.
31. 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.
32. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076.
33. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 2000.
34. Mandelbrot, B.B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **1967**, *156*, 636–638.
35. Mandelbrot, B.B. Stochastic models for the Earth's relief, the shape and the fractal dimension of the coastlines, and the number-area rule for islands. *Proc. Natl. Acad. Sci.* **1975**, *72*, 3825–3828.
36. Mandelbrot, B.B. *The Fractal Geometry of Nature*; W. H. Freeman and Company: New York, NY, USA, 1982.
37. Peitgen, H-O.; Juergens, H.; Saupe, D. *Chaos and Fractals*, 2nd ed.; Springer: New York, NY, USA, 2004.
38. Peleg, S.; Naor, J.; Hartley, R.; Avnir, D. Multiple resolution texture analysis and classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 518–523.
39. Paskas, M.P.; Gavrovskia, A.M.; Reljin, N.B. Identification of fundamental heart sounds from PCG using blanket fractal dimension. In Proceedings of the 8th Conference of the European Study Group on Cardiovascular Oscillations, Trento, Italy, 25–28 May 2014; pp. 123–124.
40. Turner, M.J.; Blackledge, J.M. Analysis of the limitations of fractal dimension texture segmentation for image characterisation. In *Fractal Geometry: Mathematical Methods, Algorithms, Applications*; Blackledge, J.M., Evans, A.K., Turner, M.J., Eds.; Horwood Publishing Limited: Chichester, England, 2002; pp. 114–137.
41. 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.
42. Que, C.-L.; Kolmaga, C.; Durand, L.-G.; Kelly, S.M.; Macklem, P.T. Phonspirometry for noninvasive measurement of ventilation: Methodology and preliminary results. *J. Appl. Physiol.* **2002**, *93*, 1515–1526.

Article

A Novel Time-Varying Spectral Filtering Algorithm for Reconstruction of Motion Artifact Corrupted Heart Rate Signals during Intense Physical Activities Using a Wearable Photoplethysmogram Sensor

S. M. A. Salehizadeh^{1*}, D. Dao², J. Bolkhovskiy³, C. Cho⁴, Y. Mendelson⁵, K. H. Chon^{6*}

¹ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

* Corresponding authors

E-mail: smasalehizadeh@gmail.com, kchon@engr.uconn.edu

² Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

³ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

⁴ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

⁵ Department of Biomedical Engineering, Worcester Polytechnic Institution, Worcester, MA, United States of America

⁶ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

Academic Editor:

Received: / Accepted: / Published:

Abstract: Accurate estimation of heart rates from photoplethysmogram (PPG) signals during intense physical activity is a very challenging problem. This is because strenuous and high intensity exercise can result in severe motion artifacts in PPG signals, making accurate heart rate (HR) estimation difficult. In this study we investigated a novel technique to accurately reconstruct motion-corrupted PPG signals and HR based on time-varying spectral analysis. The algorithm is called **S**pectral filter **a**lgorithm for **M**otion **A**rtifacts and heart rate reconstruction (SpaMA). The idea is to calculate power spectral density of both PPG and accelerometer signals for each time shift of a windowed data segment. By comparing time-varying spectra of PPG and accelerometer data, those frequency peaks resulting from motion artifacts can be distinguished from the PPG spectrum. The SpaMA approach was applied to three

different datasets and 4 types of activities: 1) training datasets from the 2015 IEEE Signal Processing Cup Database recorded from 12 subjects while performing treadmill exercise from 1 km/h to 15 km/h; 2) test datasets from the 2015 IEEE Signal Processing Cup Database recorded from 11 subjects while performing forearm and upper arm exercise. 3) Chon Lab dataset including 10 min recordings from 10 subjects during treadmill exercise. The ECG signals from all three datasets provided the reference HRs which were used to determine the accuracy of our SpaMA algorithm. The performance of the SpaMA approach was calculated by computing the mean absolute error between the estimated HR from the PPG and the reference HR from the ECG. The average estimation errors using our method on the first, second and third datasets are 0.89, 1.93 and 1.38 beats/min respectively, while the overall error on all 33 subjects is 1.86 beats/min and the performance on only treadmill experiment datasets (22 subjects) is 1.11 beats/min. Moreover, it was found that dynamics of heart rate variability can be accurately captured using the algorithm where the mean Pearson's correlation coefficient was found to be 0.98 between the power spectral densities of the reference and the reconstructed heart rate time series. These results show that the SpaMA method has a potential for PPG-based HR monitoring in wearable devices for fitness tracking and health monitoring during intense physical activities.

Keywords: Motion Artifact; Heart Rate Monitoring; Photoplethysmography; Physical Activities; Signal Processing

1. Introduction

Over the last 20 years, heart rate monitors have become widely-used training aids for a variety of sports [1]. Some heart rate monitors use photoplethysmography (PPG) technology, as it allows the device to be small and wearable [2]. The sensors, consisting of infrared light-emitting diodes (LEDs) and photodetectors, offer a simple, reliable, low-cost means of monitoring pulse rate noninvasively, both at rest and during exercise [4]. This is why they have become the sensor of choice in smart watches. HR monitoring using PPG signals has many advantages compared to using traditional ECG sensors, such as simpler hardware implementation, lower cost, and no need for daily application of electrodes [3]. 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. The pulse oximeter signal contains not only the blood oxygen saturation and heart rate (HR) data, but also other vital physiological information [4-7]. The fluctuations of photoplethysmogram (PPG) signals contain the influences of arterial, venous, autonomic and respiratory systems on the peripheral circulation. 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 [3]. Even simple knowledge of HR patterns would provide more useful clinical information than just HR and blood oxygenation, especially in situations in which a pulse oximeter is the sole monitor available. One major example of such benefits can be seen in a study by Chong et al. which show that accurate detection of atrial fibrillation can be obtained from PPG data [8].

In addition to the acquisition of HR in response to exercise, research has recently focused on obtaining heart rate variability (HRV) information from wearable sensors including devices that use PPGs [1]. Increased HRV has been associated with lower mortality rates and is affected by both age and sex [1]. During graded exercise, the majority of studies show that HRV decreases progressively up to

moderate intensities, after which it stabilizes [9]. Although there are many promising and attractive features of using pulse oximeters for vital sign monitoring, currently they are mainly used on stationary patients. This is because motion artifacts (MAs) result in unreliable HR and SpO₂ estimation [3]. Clinicians have cited motion artifacts in pulse oximetry as the most common cause of false alarms, loss of signal, and inaccurate readings [10]. During physical activities, MA contamination in PPG signals seriously interferes with HR estimation. The MAs are mainly caused by ambient light leaking into the gap between the PPG sensor surface and skin surface. Besides, the change in blood flow due to movements is another MA source [11]. In practice MAs are difficult to remove because they do not have a predefined narrow frequency band and their spectrum often overlaps that of the desired signal [12]. 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 [13–16]. 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 motion and noise artifact reduction techniques based on the concept of blind source separation (BSS). The aim of BSS is to estimate a set of uncorrupted signals from a set of mixed signals which is assumed to contain both the clean and noisy sources [3]. Some of the popular BSS techniques are independent component analysis (ICA) [17], canonical correlation analysis (CCA) [18], principle component analysis (PCA) [19], and singular spectrum analysis (SSA) [3, 20]. Kim and Yoo [21] suggested using a basic ICA algorithm and block interleaving to remove MA. Krishnan et al. [22] later proposed using frequency-domain-based ICA. However, the key assumption in ICA, namely statistical independence or uncorrelation, does not hold in PPG signals contaminated by MA [23]. Salehizadeh et al. [3] proposed a motion artifact removal algorithm using SSA. They used SSA to decompose the corrupted segment adjacent to the clean segment and chose the SSA components in the corrupted segment that had a similar frequency range to that of the adjunct clean components. Although they reported good performance, the method cannot be applied in scenarios where the HR and SpO₂ are varying rapidly due to corruption and movement. Acceleration data are also shown to be helpful to remove MA. For example, Fukushima et al. [24] suggested a spectral subtraction technique to remove the spectrum of acceleration data from that of a PPG signal. Acceleration data can be also used to reconstruct the observation model for Kalman filtering [25] to remove MA.

Two noteworthy algorithms recently published are TROIKA and JOSS [26, 27] in which sparsity-based spectrum estimation and spectral peak tracking with verification, are used to estimate and monitor heart rate during intensive physical activity, respectively. Both approaches make use of PPG and accelerometer information to obtain an accurate estimation of heart rate while running on a treadmill. TROIKA has two extra stages of signal decomposition and reconstruction using singular spectrum analysis (SSA) and it then applies temporal difference operations on the SSA-reconstructed PPG. SSA components are compared to the accelerometer signals and those components with close frequencies to the accelerometer signals are discarded and the rest are used to reconstruct the signal. In JOSS and TROIKA, spectral peak tracking with verification aims to select the spectral peaks corresponding to HR. JOSS, which has been shown to estimate HR more accurately than TROIKA, is based on the idea that the spectra of PPG signals and simultaneous acceleration signals have some common spectrum structures, and thus formulates the spectrum estimation of these signals into a joint sparse signal recovery model using the multiple measurement vector (MMV) model. MMV is used for joint spectrum estimation based on PPG and accelerometer data, which is in contrast to the single measurement vector (SMV) model that was used in TROIKA and was based on only a single PPG signal. Although JOSS has been shown to be much more accurate than previous methods for reconstruction of heart rate from MA-contaminated PPG signals, the main disadvantage of the method is it can merely provide smoothed HR reconstruction estimations. Neither time-domain PPG signal reconstruction nor heart rate variability

analysis can be done using JOSS or TROIKA. Recently, Temko proposed an approach to HR estimation based on Wiener Filtering and the Phase Vocoder (WFPV) [28]. This work showed that WFPV on average can perform better than the JOSS algorithm. The main idea of WFPV is to estimate motion artifacts from accelerometer signals and then use a Weiner filter to attenuate the motion components in the PPG signal. Phase vocoder is also applied to overcome the limited resolution of the Fourier transform and to refine the initial dominant frequency estimation.

In this paper, a new HR and also PPG signal reconstruction approach is presented using time-varying spectral analysis. The algorithm is called SpaMA and is comprised of five distinct stages: (1) time-varying power spectral density (PSD) calculation, (2) spectral filtering, (3) Motion Artifact detection, (4) HR reconstruction and (5) signal reconstruction. The idea is to calculate a window-segmented power spectral density of both PPG and accelerometer signals in real-time to scale each estimate of the PSD by the equivalent noise bandwidth of the window [29].

The simplest way to approach the first step, the PSD calculation, would be to employ a periodogram. However, it has the drawbacks that it is an inconsistent spectrum estimator, has high variance, and has leakage effects [29]. Thus, a dominant spectral peak can lead to an estimated spectrum that contains power in frequency bands where there should be no power. However, both problems can be solved by down-sampling the raw signal and then using a sufficiently small frequency step by setting a large number of frequency points. Thus, in this study we resample the signal from the original sampling frequency to 1/4 of it and then we apply the periodogram algorithm with frequency resolution of 0.001. Next, we limit the spectrum to the heart rate frequency range of [0.5Hz – 3 Hz] and take the frequency and power information of the first three peaks in the PSD at each window and signal segment. We are assuming that the heart rate component in a typical clean (motion free) PPG signal is always the dominant frequency component in the time-varying power spectrum, thus, the highest peak of the spectrum corresponds to the HR frequency. Thus, when movement happens the dominant component can be replaced by movement components which shift the HR to the second or third peak in the spectrum. So the third phase of the SpaMA algorithm is to compare the first 3 peaks and corresponding frequencies of the PPG spectrum to the first peak and frequency of the accelerometers' spectra at each window and the idea is to choose the frequency components (out of three) that are different from the accelerometers' frequency. We are assuming that when there is coherence between a spectral peak in the PPG and the accelerometers' spectra this signifies a motion noise artifact in the PPG signal and that peak should be discarded in the HR reconstruction. After discarding these movement peaks in the spectrum, the next highest peak that is closest to the estimated HR of the previous window would be chosen at each window. By reconstructing the HR frequency at each window, simultaneously we can reconstruct the PPG signal by using the power, frequency and phase of the signal that corresponds to the HR frequency. That is, we reconstruct time-domain signals from the time-frequency domain. We will show in the Results section that the new SpaMA method not only provides PPG signal and HR reconstruction but also the potential to do heart rate variability analysis on the results. We will show that SpaMA can outperform the JOSS technique in heart rate estimation by providing less error to the reference, which yields higher accuracy.

2. Materials and Methods

The SpaMA algorithm was evaluated on three different datasets. The first two datasets were provided for the IEEE Signal Processing Cup and are publically available. The three datasets are: 1.) 12 PPG training datasets (running on treadmill) from an IEEE signal processing competition [30] which was initially used in [26, 27], 2.) 11 PPG test datasets (e.g. arm exercise) from the IEEE signal processing competition and 3.) 10 PPG recordings from the Chon lab (running on treadmill).

- (1) IEEE Signal Processing Competition Training Dataset: A single-channel PPG signal, a three-axis acceleration signal, and an ECG signal simultaneously recorded from 12 Asian male subjects

ranging in age from 18 to 35. For each subject, the PPG signal was recorded from their wrist using a pulse oximeter (PO) with green LED (wavelength: 609 nm). The acceleration signal was also recorded from their wrist using a three-axis accelerometer. Both the PO and the accelerometer were embedded in a wristband, which was comfortably worn. The ECG signal was recorded from the chest and it is used as the reference heart rate. All signals were sampled at 125 Hz.

- (2) IEEE Signal Processing Competition Test Dataset: The dataset consists of 11 five minute recordings which were collected from 19 to 58 year old subjects performing intensive arm movements (e.g. boxing). For each subject, PPG signals were recorded from their wrist using a pulse oximeter with green LEDs (wavelength: 515nm). The acceleration signal was also recorded from their wrist using a three-axis accelerometer. Both the PO and the accelerometer were embedded in a wristband. An ECG signal was recorded simultaneously from their chest using wet ECG sensors. All signals were sampled at 125 Hz and sent to a nearby computer via Bluetooth.

Table 1. PPG Datasets and Experiments Settings

Subject	Dataset	Activity Type	Pulse Oximeter Type	Subject's Age/Sex
1	1 (IEEE Cup)	Type (1)	Wrist: green LED (wavelength: 609nm)	18-38 y (All Male)
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13	2 (IEEE Cup)	Type (2)	Wrist: green LED (wavelength: 515nm)	19-58 y (9 Male, 1 Female)
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24	3 (Chon Lab)	Type (4)	Forehead: Red and Infrared LED (wavelength: 660 nm, 940nm)	26-55 y (9 Male, 1 Female)
25				
26				

27				
28				
29				
30				
31				
32				
33				

- (3) Chon Lab Dataset: This dataset was recorded in the Chon Lab from 10 healthy subjects (9 male/1 female), with ages ranging from 26 to 55. For each subject, the PPG signal was recorded from their forehead using a PO (developed in our lab) with red and infrared LED (wavelength: 660 and 940 nm). The acceleration signal was also recorded from their forehead using a three-axis accelerometer. Both the pulse oximeter and the accelerometer were embedded in a headband and the signals were sampled at 80 Hz. The ECG signal was recorded as a reference from the chest using ECG sensors, sampled at 400Hz. During data recording, subjects walked, jogged and ran on a treadmill with speeds of 3, 5 and 7 mph, respectively, for 9 min. At the end, all experimental subjects were asked to perform random arbitrary movements for 1 min.

For all three datasets, we down-sampled the data to 20 Hz since the estimation of heart rate is carried out in the frequency domain and this sampling rate is sufficiently high to obtain even heart rates that are as high as 240 beats/min or 4 Hz. Moreover, this down-sampling allows us to focus on heart rates in the lower frequencies rather than in the physiologically irrelevant higher frequency ranges. Further details of this study's databases are given in Table (1). Four types of activities were involved:

- Type (1): activity involved walking or running on a treadmill for intervals of 0.5 min-1 min-1 min-1 min-0.5 min with speeds of 1-2 km/h, 6-8 km/h, 12-15 km/h, 6-8 km/h, 12-15 km/h, 1-2 km/h, respectively. The subjects were asked to purposely move the hand with the wristband to generate motion artifacts.
- Type (2): activity included various forearm and upper arm exercise which are common arm motions (e.g. shaking hands, stretching, pushing objects, running, jumping, and push-ups).
- Type (3): activity consisted of intensive forearm and upper arm movements (e.g. boxing).
- Type (4): activity involved 1 min rest, 1 min walking (3 mph), 1 min rest, 2 min jogging (5 mph), 1 min rest, 2 min running (7 mph), 1 min rest, 1 min arbitrary movement. The ECG-based reference HR was recorded in order to assess the performance of the algorithms being tested.

In summary, the first dataset includes only Type (1), the second dataset includes both Type (1) and (2) activities, and the third dataset includes only Type (4) activities.

2.1. Methodology

The procedure for our new HR monitoring algorithm during intensive movements is presented in Table (2). Details of each stage will be described in subsections *i* to *v*.

Table 2. The proposed SpaMA algorithm: HR and PPG signal reconstruction

--

Stage 1. Time-Varying Spectral analysis

- 1.1. Down sample the PPG and Accelerometer signal to 20 Hz.
- 1.2. Compute the power spectral density of both PPG and Accelerometers [0-10 Hz].

Stage 2. Spectral Filtering

- 2.1. Assume HR to be in the frequency range of [0.5Hz – 3Hz], this accounts for both low and high heart rates.
- 2.2. The first highest three peaks and their corresponding frequencies in the PPG filtered spectrum are assumed to have HR information.
- 2.3. Only the largest frequency peak of the accelerometers' spectra is used for MA detection in stage 3.

Stage 3. Motion Artifact Detection

- 3.1. Compare the frequencies of the three peaks in the PPG spectrum with the frequency of the largest peak in the accelerometers' spectra. If the first or second largest peaks in the PPG spectrum are similar to that of the accelerometers' peaks, then motion artifact is present in the PPG.
- 3.2. If motion artifact is detected from 3.1, then the corresponding frequency peak (usually the first or second largest peak) in the PPG spectrum should be discarded.

Stage 4. Heart Rate Tracking and Extraction from PPG Spectrum

Case (1): From 3.1- if the spectrum is corrupted by movement and only the first largest peak is corrupted, then the HR frequency should be the frequency of the second peak in the spectrum.

Case (2): From 3.1- if the spectrum is corrupted by movement and both the first and second largest peaks have similar frequencies to those of the accelerometers' peaks, then the HR frequency should be the frequency of the third peak in the spectrum.

Case (3): Due to a gap between the pulse oximeter and a subject's skin, the HR frequency cannot be extracted from the spectrum and in this case the previous HR frequency is used or for offline implementation a cubic spline interpolation can be applied to fill in the missing HR information.

Stage 5. PPG Signal Reconstruction

- 6.1. The PPG signal is reconstructed by using the amplitude, frequency and phase information corresponding to the HR components (extracted in stage 4) that are calculated from the spectrum at each window.

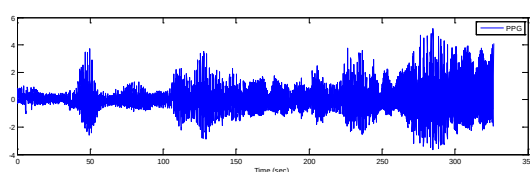
❖ Heart Rate Variability Analysis

By using a sample-by-sample windowing strategy, HR can be extracted, from which dynamics of heart rate variability analysis can be obtained on the motion artifact-removed reconstructed HR time series.

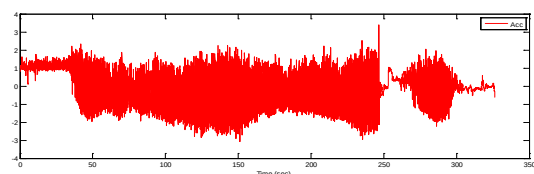
i. Time-Varying Spectral Analysis of PPG and Accelerometer Data

We produce a time-varying spectrum by taking a T-sec window of the signal and computing the power spectral density (PSD) of the segment and then sliding the window through the whole dataset which yields a time-frequency matrix in which each array represents the power of the signal corresponding to a specific frequency and sliding time-step (shift) of S-sec. The sliding process and frequency step specify the resolution and dimension of the time-frequency matrix. In this study, we take two different sliding window approaches depending on the application. For estimating either heart rates or heart rate variability, data are shifted sample-by-sample with no overlap for the entire dataset. This is because we are interested in capturing beat-to-beat dynamics of HRV which requires sample-to-sample estimation of PSD. Given our down-sampled data to 20 Hz in some of the database, each data point is shifted by 0.05 seconds. For estimating only the heart rates, we shift the data segment-by-segment rather than sample-by-sample. This coarse-grain windowing approach has less computational cost and it can provide good tracking of heart rates, but it cannot be used for HRV. The window segment length T was set to 8 seconds and was shifted by 2 seconds. We chose an 8 second data segment and a shift of 2 seconds because one of the goals of this work is to compare our algorithm's results to other algorithms compared in this work (TROIKA, JOSS and WFPV) which have used this chosen data segment length and time shift [26, 27]. Moreover, the assumption of 8 second data length largely stems from the fact that heart rates do not change instantaneously, hence, an 8 second duration is a reasonable choice.

As a representative example, the resultant frequency components in the time-frequency matrix of recordings from subject #8 from the competition training dataset, for a window length of 8 seconds that is shifted by every 2 seconds, is shown in Fig. 1.



(a)



(b)

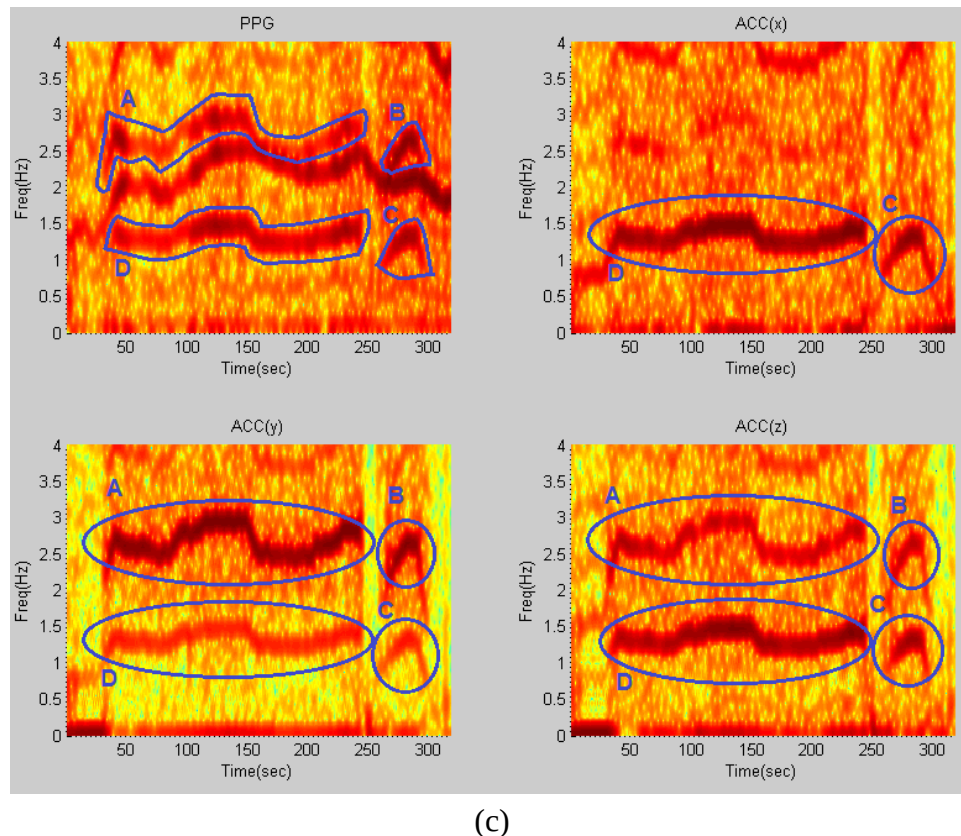


Figure 1. Time-Frequency spectra of recording #8 from dataset (1): **(a)** PPG signal, **(b)** simultaneous Accelerometer-Z signal, **(c)** (Top-Left) TF spectrum of PPG, (Top-Right) TF spectrum of ACC(x), (Bottom-Left) TF spectrum of ACC(y), (Bottom-Right) TF spectrum of ACC(z); all computed from stage (1) of the algorithm. Blue circles and letters represent movement elements in all four spectra.

The panels of Fig. 1a and 1b show a PPG time series and the z-axis accelerometer data, respectively. From the upper left panel of Fig. 1c, which represents the time-frequency plot of the PPG signal, it is observed that there are three dominant frequency components - one of them represents HR and the other two are similar to those of the accelerometers' spectra shown in the upper right and lower left and right panels of Fig. 1c. This figure illustrates 4 motion artifact elements (A, B, C, D) that are present in exactly the same areas among all spectra. By comparing the time-frequency (TF) spectrum of PPG to those of the accelerometers' spectra, we can detect that the marked dynamics (A, B, C and D) in the PPG spectrum shares the same frequency dynamics as those of the accelerometer spectra marked in circles. Hence, both the top and bottom marked lines in the PPG spectrum most likely represent the motion artifacts, and the unmarked frequency represents the HR. The next section details how these motion artifact frequency dynamics are detected and filtered.

ii. Spectral Filtering

After obtaining the power spectral density at each window, HR frequency is assumed to be confined in the range [0.5 Hz – 3 Hz], which takes into account both at rest and high HR due to either tachycardia or exercise scenarios. Hence, for HR estimation, the strategy is to eliminate frequencies that are outside of this HR frequency range as they are most likely due to motion artifacts or harmonics of the HR frequency.

In general, HR frequency in the power spectral density of PPG at each window can have three different scenarios: (1) PPG is devoid of MA and there is no spatial gap between the sensor and the subject's skin during recording, (2) PPG is corrupted by MA and there is no spatial gap between the sensor and the subject's skin during recording and (3) There is a spatial gap between the sensor and the subject's skin during recording. For the ideal case (1), HR can be extracted and it is most likely represented as the highest peak in the PPG spectrum. For case (2), MA dynamics can result in predominately either one or two dominant peaks depending on the severity of repetitive motions, and the HR peak is relegated to either the second or third highest peak. Non-repetitive motion artifacts will show up as a broadband spectrum without a dominant peak if they are not severe [31]. The only scenario that makes it difficult to extract HR from the spectrum is scenario (3) when there is a spatial gap between the PPG sensor and the subject's skin during recording. In this scenario, assuming that the motion artifacts are short lasting, the missing HR values can be interpolated using the cubic spline approach.

Fig. 2 shows a representative filtered time-frequency spectral plot of a PPG signal. This step in the SpaMA process involves retaining only the three largest frequency peaks at each time point within the defined HR range (30–180 bpm) and they are represented as blue, green and red colors, respectively. It is our opinion that retaining only the three largest frequency peaks at each time point is reasonable for the first two cases as outlined above.

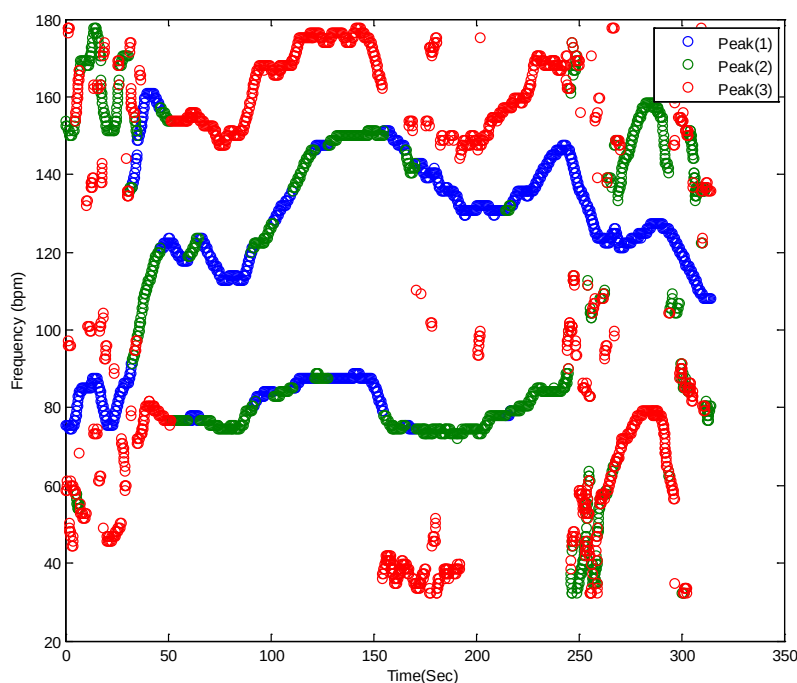


Figure 2. Spectral Filtering. PPG time-frequency spectrum: Blue, Green and Red circles correspond to the first three highest peaks in the defined HR frequency range of (30–180 bpm), respectively, at each sliding window.

iii. Motion Artifact Detection

Fig. 3a illustrates a PPG spectrum which is identical to Fig. 2, but it also identifies the frequencies associated with accelerometers, as marked by the shaded areas and the letters A–D, in the top left and two bottom panels of Fig. 1c. By removing the accelerometers' related frequencies in Fig. 3a, the

remaining frequency dynamics which should represent HR frequency and its harmonics are shown in Fig. 3b.

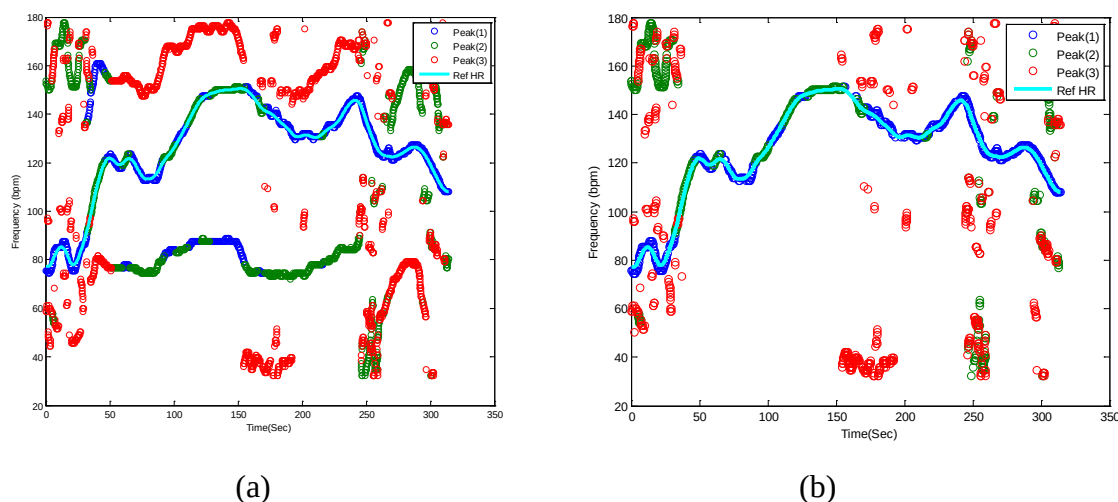


Figure 3. Motion Artifact Detection in the PPG spectrum. **(a)** Filtered PPG spectrum with movement and HR components: Shaded yellow elements (A, B, C, D) represent motion frequency components in the PPG spectrum, and the light blue line is the reference HR from clean reference ECG signal, **(b)** Filtered PPG spectrum after removing motion artifact frequency components.

iv. Heart Rate Tracking & Extraction

The next step is to identify HR frequencies with time from Fig. 3b. Note that in Fig. 3b, there are three peaks at each time instance, thus, the question is how to identify which of the three peaks represents the HR at each time point. For the initial time window of 8 seconds, we require a clean data segment so that true HR can be determined. This scenario is case (1) described above in the spectral filtering section, and the detection of HR is simply the highest peak in the spectrum. The next step is to estimate HR for each sliding window of data. At this step of the algorithm, the goal is to choose a HR peak in the PPG spectrum with the knowledge of estimated HR values in previous time windows. In this step there are two main scenarios: (1) no peak exists in the spectrum that can represent HR, and (2) there is a spectral peak among the first three highest peaks of spectrum that belongs to the HR component. In case (1), where HR is not detectable in the window (e.g. due to spatial gap between the PO sensor and skin), in real-time implementation the algorithm takes the previous window's HR value as the current HR (or simply uses the moving average of several past HR beats or some other variant), however in offline processing, a cubic spline interpolation can be used to fill in the missing HR information. In the more general case (2), where the HR peak is among the first three highest peaks in the spectrum, three possible scenarios can occur: (2-A) the windowed PPG signal is clean and the first highest peak in the spectrum represents the HR fundamental frequency, (2-B) the windowed PPG signal is corrupted by movement and at most two of the spectral peaks represent the accelerometers' frequency components, thus the second or the third peak corresponds to HR, (2-C) while the HR spectral peak is detectable, the difference between its value and that of the previous HR is more than 10 bpm, so it will be replaced by the most recent HR value from a previous window segment (or a moving average of several past HR beats or some other variant). We set a criterion that the HR value cannot change more than 10 BPM from a previous time window. In Fig. 3b, these cases are illustrated. For example, in most cases, the blue circle which represents the largest spectral peak is chosen but in other cases, either green or red circles are chosen for certain time points. For the HR peaks associated with either the green or red circles, they are chosen because either the first two highest peaks are related to accelerometers or the highest magnitude peak deviates more than 10 BPM from the previous HR value. Fig. 4 summarizes the flowchart of HR tracking and extraction procedures.

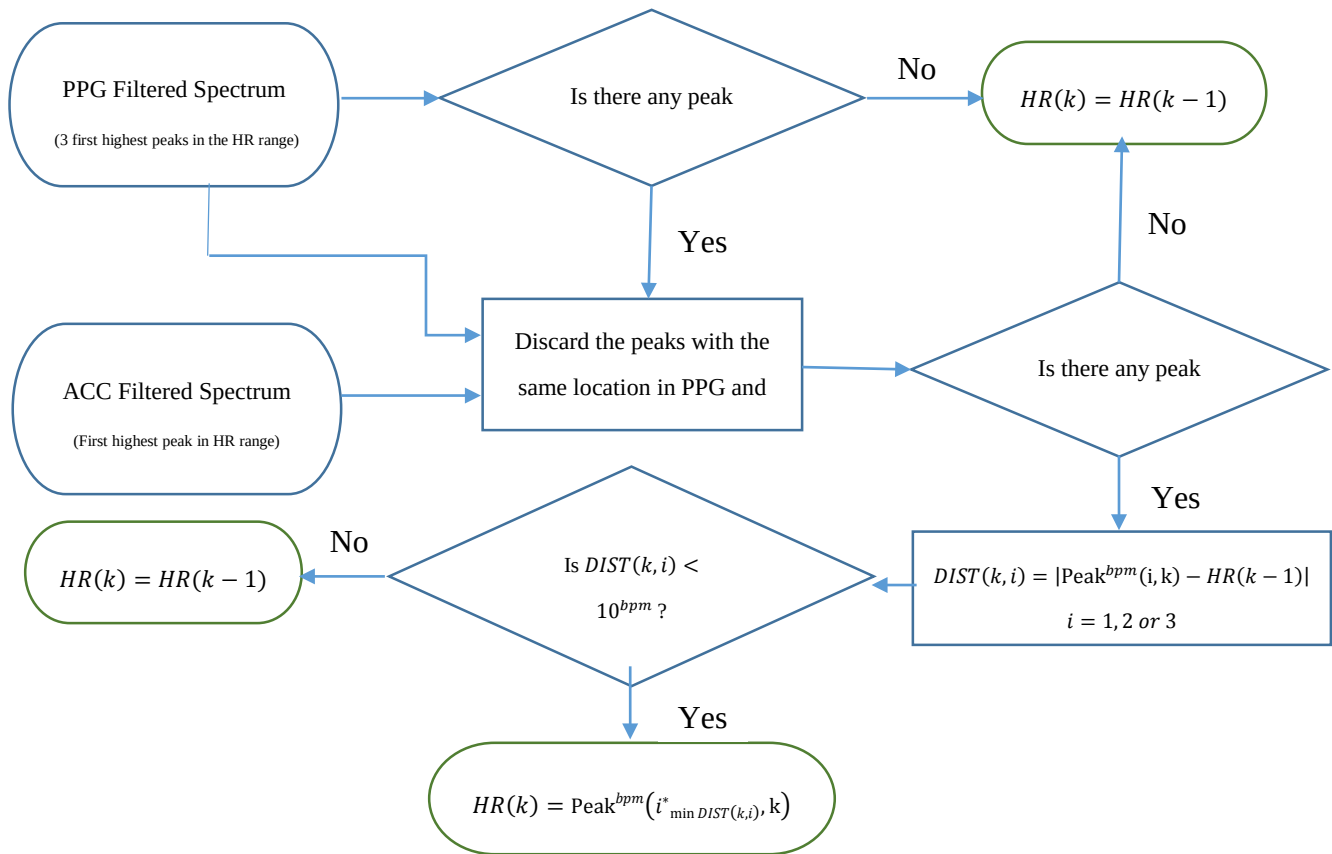


Figure 4. Flowchart of HR Tracking and Extraction

Fig.5 shows the extracted HR (red color) from PPG spectra of recording#8 from the competition training dataset using our proposed approach along with the reference ECG-derived HR (black color). In order to calculate the performance of the SpaMA algorithm, the error value in each time window was calculated from the estimated HR to the reference ECG-derived HR. Two measurement indices of absolute error similar to the indices in [22] were used.

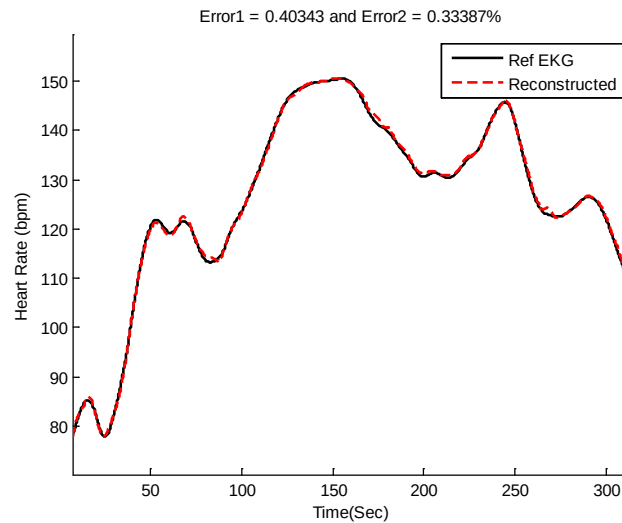


Figure 5. Comparison of reconstructed HR obtained from SpaMA to reference HR obtained from simultaneous ECG recordings.

$$Error(1) = \frac{1}{W} \sum_{k=1}^w |HR_{SpaMA}(k) - HR_{ref}(k)| \quad (1)$$

$$Error(2) = \frac{1}{W} \sum_{k=1}^W \frac{|HR_{SpaMA}(k) - HR_{ref}(k)|}{HR_{ref}(k)} \times 100\% \quad (2)$$

where W is the total number of windows.

v. PPG Signal Reconstruction for HRV analysis

For HRV analysis application, the above-described procedures are identical but the only difference is the beat-by-beat shift of data rather than the 8 second data segment shift or its variant. The PPG signal can now be reconstructed using heart rate frequency, amplitude and phase changes, window-by-window using the sample-by-sample windowing process:

$$Rec_{Signal}(k) = A_{HR}(k) \times \sin(2\pi t(k)f_{HR}(k) + \varphi_{HR}(k)) \quad (3)$$

where $k = 1, \dots, N$ and N is number of signal samples and total number of windows. $A_{HR}(k)$ and $\varphi_{HR}(k)$ are calculated according to the power of the signal for HR frequencies in the PSD and phase angles of complex elements in the FFT matrix that correspond to HR frequencies. The left and right panels of Fig. 6 show the reconstructed PPG versus the original PPG and their HRV time series, respectively. Fig. 7 shows comparison of HRV spectra between the reference and the reconstructed HRV time series (e.g., as shown in Fig. 6) from the MA-contaminated PPG signal for dataset #8. Note that for computing HRV, we are not concerned about matching the amplitude of the reference HR, as we are interested only in the dynamics of the fluctuations in the heart rates.

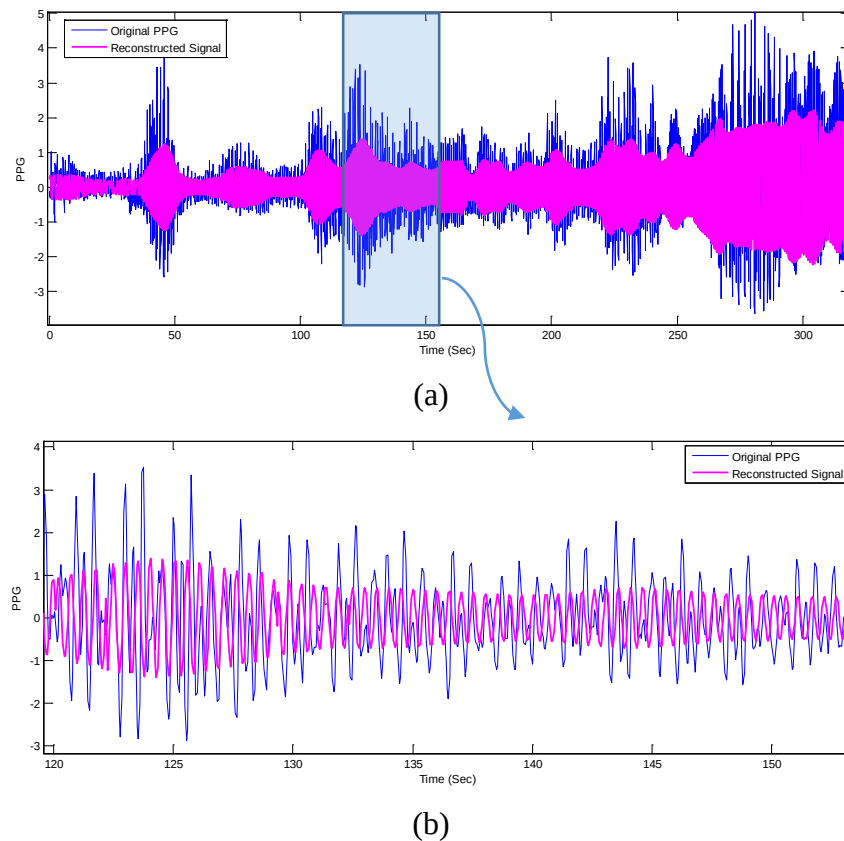


Figure 6. PPG signal reconstruction. **(a)** Comparison between reconstructed PPG and original recording #8 from IEEE competition training dataset, **(b)** Zoomed-in version of (a)

Given that we are able to estimate quite accurate heart rates, it is not surprising to observe similar frequency dynamics between the reference and reconstructed HRV time series.

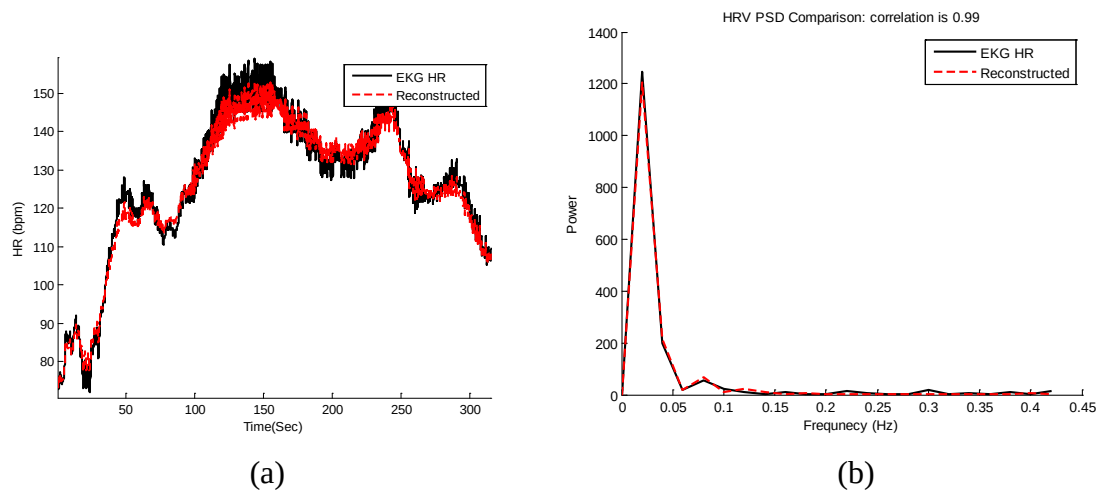


Figure 7. Heart Rate Variability Analysis **(a)** Time-domain comparison of reconstructed and reference HR. **(b)** Spectral comparison of heart rate variability between reconstructed HR and reference HR calculated from the reference ECG using Pan & Tompkins peak detection approach [32].

3. Results

Table 3 represents the average absolute error (E1) and the average absolute error percentage (E2) of the proposed SpaMA algorithm on all 3 datasets, respectively. Our SpaMA algorithm is compared to three recently-developed algorithms: TROIKA, JOSS and WFPV. The results in Table 3 show that SpaMA has better performance than JOSS and TROIKA for all 12 subjects in the first datasets. In comparison to WFPV, the proposed SpaMA approach outperforms WFPV on average across all 23 subjects in both datasets (1) and (2). The total average of E1 of SpaMA is less than 2 beats per minutes for all 33 subjects. The average of E1 across the treadmill experiment recordings (activity Type 1- IEEE dataset and Type 4- Chon Lab dataset) is around 1 beat per minute for all 22 subjects.

Figs. 8-10 show the reconstructed HR and corresponding PSD of a sample-sample windowed HR in comparison to the reference HR from ECG. The results for recording #9 from the first dataset (IEEE Competition Training dataset) and activity Type 1 (e.g. running on treadmill) are shown in Fig.8. We can see that the E1 for this particular subject is as low as 0.4 bpm and the correlation between the PSD of reconstructed HR and reference HR is as high as 96%.

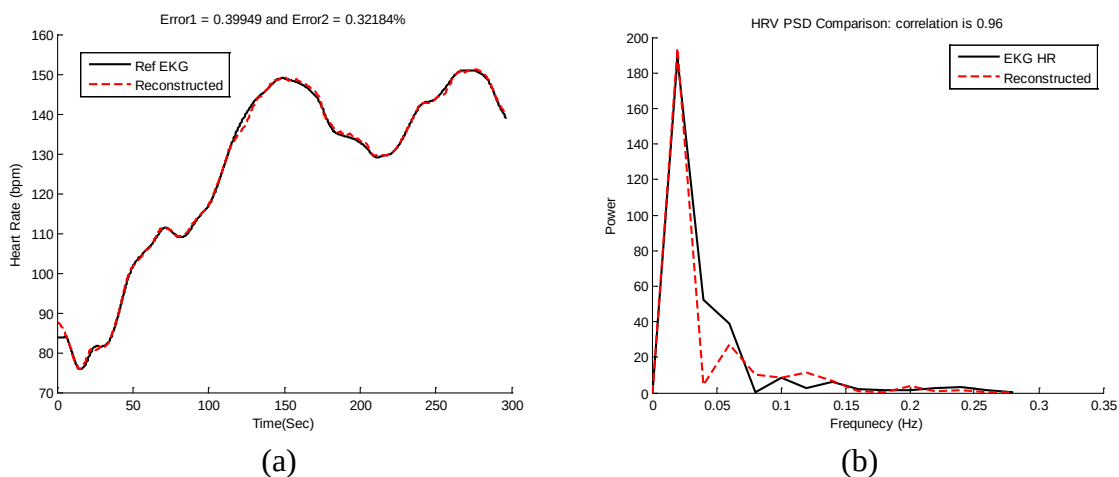


Figure 8. Subject 9 (IEEE Competition Training Dataset). **(a)** Reconstructed HR vs. reference HR. **(b)** Spectral comparison of reconstructed HR and reference HR (estimated from reference ECG)

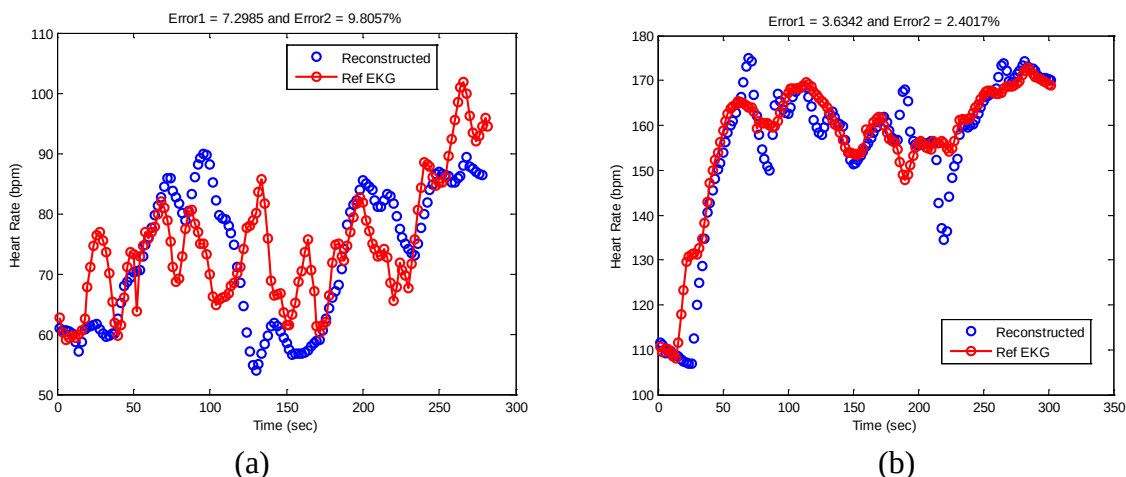


Figure 9. Reconstructed HR vs. reference HR: **(a)** Subject 14 (IEEE Competition Test Dataset). **(b)** Subject 16 (IEEE Competition Test Dataset).

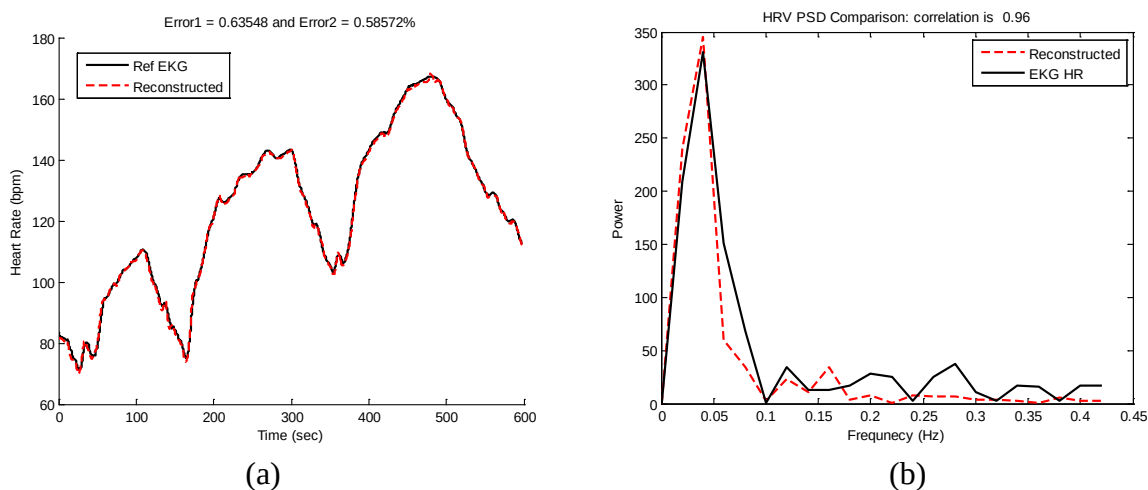


Figure 10. Subject 30 (Chon Lab Dataset). **(a)** Reconstructed HR vs. reference HR. **(b)** Spectral comparison of reconstructed HR and reference HR (estimated from reference ECG)

Fig. 9a illustrates the comparison between the reconstructed HR and the reference HR for subject #14 which has the highest errors. Subject #14 belongs to the second dataset (IEEE Competition Test dataset) and Type 2 activities (e.g. jumping). It can be seen that the largest error is obtained when both the physiological HR and the motion artifacts change rapidly so that the true HR cannot be reliably estimated. Fig. 9b shows the comparison results of reconstructed HR and reference HR for recording #16 of the second dataset (IEEE Competition Test dataset) which has the Type 3 activities (e.g. arm exercise).

The results for recording #30 from the third datasets (Chon Lab datasets) and activity Type 4 (e.g. running on treadmill) are shown in Fig. 10. It can be seen that the E1 for this subject is around 0.6 bpm and the correlation between the PSD of reconstructed HR and reference HR is as high as 99% for LF and 0.96 for HF frequency range. All subjects' results are provided in Table 4.

Table 3. SpaMA Algorithm Performance Comparison

Subject	Dataset	Activity Type	TROIKA		JOSS		WFPV		SpaMA	
	1 (IEEE Cup)	Type (1)	E1	E2%	E1	E2%	E1	E2%	E1	E2%
1			2.87	2.18	1.33	1.19	1.23	-	1.23	1.14
2			2.75	2.37	1.75	1.66	1.26	-	1.59	1.30
3			1.91	1.50	1.47	1.27	0.72	-	0.57	0.45
4			2.25	2.00	1.48	1.41	0.98	-	0.44	0.31
5			1.69	1.22	0.69	0.51	0.75	-	0.47	0.31
6			3.16	2.51	1.32	1.09	0.91	-	0.61	0.45
7			1.72	1.27	0.71	0.54	0.67	-	0.54	0.40
8			1.83	1.47	0.56	0.47	0.91	-	0.40	0.33
9			1.58	1.28	0.49	0.41	0.54	-	0.40	0.32
10			4.00	2.49	3.81	2.43	2.61	-	2.63	1.59
11			1.96	1.29	0.78	0.51	0.94	-	0.64	0.42
12			3.33	2.30	1.04	0.81	0.98	-	1.20	0.86
mean±std			2.42±0.8	1.82±0.5	1.28±0.9	1.02±0.6	1.04±0.5	-	0.89±0.6	0.65±0.4
13	2 (IEEE Cup)	Type (2)					3.58	-	3.41	4.25
14							9.66	-	7.29	9.80
15							2.31	-	2.73	2.21
16		Type (3)					4.93	-	3.18	2.11
17							3.07	-	3.01	2.52
18							2.67	-	4.46	3.23
19							3.11	-	3.58	3.98
20		Type (2)					2.10	-	1.94	1.66
21		Type (3)					3.22	-	2.56	2.02
22							4.35	-	3.12	3.28
23		Type (2)					0.75	-	1.72	1.97
mean±std							3.61±2.2	-	3.36±1.5	3.33±2.2
Type (1,2)							2.27±2.0	-	1.93±2.0	2.07±1.7
mean±std										
24	3 (Chon Lab)	Type (4)							0.88	0.91
25									1.03	0.83
26									1.10	0.90
27									1.64	1.54
28									1.41	1.12
29									0.82	0.70
30									0.63	0.58
31									4.78	3.87

32				0.95	0.79
33				0.62	0.52
mean±std				1.38±1.2	1.17±1.0
Total: mean±std				1.86±1.6	1.70±1.8

Table 4 represents the correlation and statistical difference using the student's t-test between PSD of estimated and reference HRV in both LF (0.04–0.15 Hz) and HF (0.15–0.4 Hz) frequency ranges. The correlation values in the table are calculated based on Pearson's linear correlation coefficient. As shown in Table 4, there was no difference between the reference and our derived HRV for LF and the difference was seen in only 4 out of 10 subjects for HF. Table 5 shows some of the widely-reported time-domain HRV parameters such as the mean HR, standard-deviation (SDNN) of the normal-to-normal (NN) interval, root-mean-square of successive difference (RMSSD) of the NN interval, and the number of interval differences of successive NN intervals greater than 50 ms divided by the total number of NN intervals (pNN50) estimated from SpaMA in comparison to the reference ECG NN interval. None of these parameters were found to be significantly different between our algorithm-derived and the reference HRV.

Table 4. Frequency Domain HRV analysis Comparison: PSD of SpaMA vs. reference

Subjects	Correlation	
	LF ¹	HF
1	0.99	0.98
2	0.99	0.96
3	0.99	0.95* ²
4	1.00	0.99
5	1.00	0.99
6	0.99	0.96*
7	0.98	0.92*
8	0.97	0.90*
9	1.00	0.99
10	1.00	0.99
mean	0.99	0.96

Table 5. Time Domain HRV analysis Comparison: SpaMA vs. reference HRV

Subjects	SDNN		meanNN		RMSSD		pNN50	
	SpaMA	Reference	SpaMA	Reference	SpaMA	Reference	SpaMA	Reference
1	2620.75	2566.47	10481.89	10480.72	33.24	18.05	0.001	0.020
2	2115.44	2079.58	9908.00	10020.00	25.93	16.32	0.011	0.019
3	3173.73	3177.68	10764.20	10829.06	89.70	56.15	0.019	0.207
4	2517.78	2533.20	10376.95	10426.26	13.54	19.58	0.001	0.030

¹ LF is [0.04–0.15] Hz and HF is [0.15–0.4] Hz

² (*) indicates significantly different (P-value>0.05)

5	2654.42	2670.32	10846.04	10990.08	11.88	18.59	0.003	0.018
6	2012.53	1974.65	9737.35	9827.63	39.64	21.17	0.004	0.025
7	3056.36	2925.19	12519.74	13134.05	27.66	30.61	0.015	0.071
8	3133.76	2756.66	10504.00	10530.00	32.57	36.38	0.002	0.003
9	2195.08	2142.53	10499.81	10470.06	8.23	13.01	0.002	0.004
10	2454.57	2406.96	12936.62	12981.21	41.52	20.28	0.006	0.024
p-value	> 0.05		> 0.05		> 0.05		> 0.05	

4. Discussion and Conclusions

In this study, a new approach (SpaMA) based on time-varying spectral analysis of the PPG signal is introduced to address Heart Rate monitoring in the real world with challenges ranging from a subject who makes sudden movements but is otherwise sedate, to intensive physical activities. The idea behind the proposed SpaMA approach is to compare spectral changes in PPG and accelerometer signals. Three different datasets have been used to verify the algorithm performance. Each dataset reflects different types of activities and movements. In all of the experiments, the reference HR was calculated from an ECG signal that was collected simultaneously with the PPG signal. The estimated HR was calculated from the spectrum of PPG in 8 second time windows. It has been shown in the results section that the proposed SpaMA algorithm can be used for tracking HR changes during severe motion artifacts with an average error of just 1.86 bpm compared to that of the reference ECG; these results are superior to three other algorithms tested: TROIKA, JOSS and WFPV [26, 27].

Out of 33 recordings, 23 are from a wrist pulse oximeter, and the rest of the data were recorded by a forehead pulse oximeter. The results from Table 3 show that the SpaMA algorithm can be applied to monitor HR from both wrist and forehead pulse oximeters. By comparing the performance of the algorithm for treadmill experiments (dataset 1 and dataset 3), the error is lower by almost 1 beat using a wrist pulse oximeter. However, we cannot conclude from this result that the wrist PPG provides less error than the forehead, as the experiments used different subjects and were two separate studies. The prior algorithms (TROIKA, JOSS and WFPV) were tested using only the wrist-based PPG signals as the inventors of these algorithms did not have access to forehead PPG sensors. Our algorithm, tested on data from both PPG sensor locations, proved to be effective regardless of the location of the PPG sensor.

We made several observations while analyzing the data. The tracking ability of the SpaMA algorithm decreased as the frequency changes during recordings increased. This phenomenon mostly was observed while dealing with the second set of datasets from the IEEE competition, which involved Type (2) and Type (3) activities. These types of exercises involved more abrupt movements which consequently made it more difficult to track the HR-related frequencies in the spectrum. In the three datasets that have been analyzed, recordings #10 and 14 are examples of this phenomenon.

The strength of the PPG's LED is one of the most important factors determining the SpaMA algorithm performance. Movement induces much less spectrum corruption (shift) in the PPG if the LED is sufficiently strong. A reduction in the strength of the PPG signal can also be caused by ambient light leaking into the gap between a PPG sensor and the skin surface [27]. This is because the power of the signal is dependent on the depth and reflection of the light from the pulse oximeter to the subject's skin. This gap between skin and the planar substrate where the LEDs and PD are mounted may be the result of movement during physical activities or the shape of tissue that the sensors touch. Among the three datasets, low performance for recordings #16 and 31 is the result of a weak PPG signal most likely due to a gap between the sensor and skin caused by motion artifacts.

By using the sample-by-sample windowing process, the proposed SpaMA [33] algorithm can be utilized for both Heart Rate monitoring and HRV analysis in both frequency- and time-domains. From the Results section, it can be observed that the algorithm is able to replicate both the low frequency (0.04-0.15Hz) and the high frequency (0.15-0.4) dynamics well, albeit better the former than the latter, when compared to the reference HRV. For time-domain HRV measures, the mean HR, SDNN, RMSSD, and pNN50 from our algorithm were all found to be not significantly different than the reference HRV. It has long been shown that during dynamic exercise, heart rate increases due to both a parasympathetic withdrawal and an augmented sympathetic activity [34, 35]. The relative role of the two drives depends on the exercise intensity [33, 36, 37]. Analysis of HRV permits insight into this control mechanism [38]. Also, being able to do HRV analysis from PPG even during movement and physical activities would be an advantage for detecting and diagnosing many cardiovascular diseases using only PPG recordings.

The proposed SpaMA algorithm can be potentially implemented in real time. We have found that the algorithm written in Matlab, version R2014 and CPU 3.4 GHz takes only 110 msec per 8 second segments. Therefore, given the high accuracy of the proposed approach in estimating HR despite severe motion artifacts, this method has the potential to be applicable for implementation on wearable devices such as smart watches and PPG-based fitness sensors.

Acknowledgments

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

Author Contributions

S. M. A. Salehizadeh: contributed in development of SpaMA algorithm, analysis of its performance and also participated in data collection procedure.

D. Dao: contributed in designing the experiment and data collection.

J. Bolkhovsky: contributed in designing the experiment and data collection

C. Cho: contributed in implementing the algorithm as a smart phone application which is ongoing.

Y. Mendelson: contributed as a research advisor.

K. H. Chon: contributed as a research advisor.

Conflicts of Interest

“The authors declare no conflict of interest”.

References

1. Achten, J. and A.E. Jeukendrup, *Heart rate monitoring: applications and limitations*. Sports medicine (Auckland, N.Z.), 2003. **33**(7): p. 517-538.
2. Yilmaz, T., R. Foster, and Y. Hao, *Detecting Vital Signs with Wearable Wireless Sensors*. Sensors, 2010. **10**(12): p. 10837.

3. Salehizadeh, S.M.A., et al., *Photoplethysmograph Signal Reconstruction based on a Novel Motion Artifact Detection-Reduction Approach. Part II: Motion and Noise Artifact Removal*. Annals of Biomedical Engineering, 2014. **42**(11): p. 2251-2263.
4. Chon, K.H., S. Dash, and J. Kihwan, *Estimation of Respiratory Rate From Photoplethysmogram Data Using Time–Frequency Spectral Estimation*. Biomedical Engineering, IEEE Transactions on, 2009. **56**(8): p. 2054-2063.
5. Lee, J. and K.H. Chon, *Respiratory Rate Extraction Via an Autoregressive Model Using the Optimal Parameter Search Criterion*. Annals of Biomedical Engineering, 2010. **38**(10): p. 3218-3225.
6. Dash, S., et al., *Estimation of Respiratory Rate From ECG, Photoplethysmogram, and Piezoelectric Pulse Transducer Signals: A Comparative Study of Time–Frequency Methods*. Biomedical Engineering, IEEE Transactions on, 2010. **57**(5): p. 1099-1107.
7. Jinseok, L. and K.H. Chon, *Time-Varying Autoregressive Model-Based Multiple Modes Particle Filtering Algorithm for Respiratory Rate Extraction From Pulse Oximeter*. Biomedical Engineering, IEEE Transactions on, 2011. **58**(3): p. 790-794.
8. Jo Woon, C., et al., *Arrhythmia Discrimination Using a Smart Phone*. Biomedical and Health Informatics, IEEE Journal of, 2015. **19**(3): p. 815-824.
9. Laughlin, M.H., *Cardiovascular response to exercise*. Am J Physiol, 1999. **277**(6 Pt 2): p. S244-59.
10. Jubran, A., *Pulse oximetry*. Crit Care, 1999. **3**(2): p. R11-r17.
11. Maeda, Y., M. Sekine, and T. Tamura, *Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography*. J Med Syst, 2011. **35**(5): p. 969-76.
12. Thakor, N.V. and Y.S. Zhu, *Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection*. IEEE Trans Biomed Eng, 1991. **38**(8): p. 785-94.
13. Diniz, P., *Adaptive filtering: algorithms and practical implementation*. 2008: . Springer Science, Business Media L.L.C., .
14. Kalman, R.E., *A New Approach to Linear Filtering and Prediction Problems* Transactions of the ASME – Journal of Basic Engineering, 1960. **Series D**(82).
15. Morbidi, F., et al., *Application of Kalman Filter to Remove TMS-Induced Artifacts from EEG Recordings*. Control Systems Technology, IEEE Transactions on, 2008. **16**(6): p. 1360-1366.
16. Seyedtabaai, S.S.a.L., *Kalman filter based adaptive reduction of motion artifact from photoplethysmographic signal*. World Academy of Science, Engineering and Technology, 2008. **37**.
17. Comon, P., *Independent component analysis, A new concept?* Signal Processing, 1994. **36**(3): p. 287-314.
18. Thompson, B., *Canonical Correlation Analysis: Uses and Interpretation*. 1984: SAGE Publications.
19. Jolliffe, I.T., *Principal Component Analysis*. 2 ed. Springer Series in Statistics. 2002: Springer-Verlag New York. 488.
20. Elsner, J.B. and A.A. Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. 1996: Springer.
21. Kim, B.S. and S.K. Yoo, *Motion artifact reduction in photoplethysmography using independent component analysis*. IEEE Trans Biomed Eng, 2006. **53**(3): p. 566-8.
22. Krishnan, R., B. Natarajan, and S. Warren, *Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data*. Biomedical Engineering, IEEE Transactions on, 2010. **57**(8): p. 1867-1876.
23. Yao, J. and S. Warren, *A short study to assess the potential of independent component analysis for motion artifact separation in wearable pulse oximeter signals*. Conf Proc IEEE Eng Med Biol Soc, 2005. **4**: p. 3585-8.
24. Fukushima, H., et al., *Estimating heart rate using wrist-type Photoplethysmography and acceleration sensor while running*. Conf Proc IEEE Eng Med Biol Soc, 2012. **2012**: p. 2901-4.

25. Boreom Lee, J.H., Hyun Jae Baek, Jae Hyuk Shin, Kwang Suk Park and Won Jin Yi, *Improved elimination of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry*. Physiological Measurement, 2010. **31**(12): p. 1585.
26. Zhilin, Z., P. Zhouyue, and L. Benyuan, *TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals During Intensive Physical Exercise*. Biomedical Engineering, IEEE Transactions on, 2015. **62**(2): p. 522-531.
27. Zhang, Z., *Photoplethysmography-Based Heart Rate Monitoring in Physical Activities via Joint Sparse Spectrum Reconstruction*. Biomedical Engineering, IEEE Transactions on, 2015. **PP**(99): p. 1-1.
28. Temko, A. *Estimation of Heart Rate from Photoplethysmography during Physical Exercise using Wiener Filtering and the Phase Vocoder*. in *IEEE EMBS International Conference on Engineering in Medicine and Biology*. 2015. Milan, Italy: IEEE.
29. Stoica, P. and R.L. Moses, *Introduction to Spectral Analysis*. 1997: Prentice Hall.
30. ; Available from: <http://www.signalprocessingsociety.org/spcup2015/index.html>.
31. Marmarelis, V.Z., et al., *Nonlinear analysis of renal autoregulation under broadband forcing conditions*. Ann Biomed Eng, 1993. **21**(6): p. 591-603.
32. Pan J, T.W., *A real-time QRS detection algorithm*. IEEE Trans Biomed Eng, 1985. **32**(3): p. 230-236.
33. Perini, R., et al., *The influence of exercise intensity on the power spectrum of heart rate variability*. Eur J Appl Physiol Occup Physiol, 1990. **61**(1-2): p. 143-8.
34. Iellamo, F., *Neural mechanisms of cardiovascular regulation during exercise*. Auton Neurosci, 2001. **90**(1-2): p. 66-75.
35. Bernardi, L. and M.F. Piepoli, *[Autonomic nervous system adaptation during physical exercise]*. Ital Heart J Suppl, 2001. **2**(8): p. 831-9.
36. Sarmiento, S., et al., *Heart rate variability during high-intensity exercise*. Journal of Systems Science and Complexity, 2013. **26**(1): p. 104-116.
37. Roure, R., et al., *Autonomic nervous system responses correlate with mental rehearsal in volleyball training*. Eur J Appl Physiol Occup Physiol, 1998. **78**(2): p. 99-108.
38. Aubert, A.E., B. Seps, and F. Beckers, *Heart rate variability in athletes*. Sports Med, 2003. **33**(12): p. 889-919.

A Robust Motion Artifact Detection Algorithm for Photoplethysmographic Signals using Time-Frequency Spectral Features

S. M. A. Salehizadeh^{1*}, D. Dao^{1*}, J. W. Chong², C. Cho³, D. McManus⁴, C. E. Darling⁵, Y. Mendelson⁶, K. H. Chon⁷

¹ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

*Corresponding authors

E-mail: smasalehizadeh@gmail.com, dkdao2013@gmail.com

² Department of Biomedical Engineering, Worcester Polytechnic Institution, Worcester, MA, United States of America

³ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

⁴ Department of Department of Medicine and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States of America

⁵ Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, United States of America

⁶ Department of Biomedical Engineering, Worcester Polytechnic Institution, Worcester, MA, United States of America

⁷ Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States of America

Abstract— Motion and noise artifacts (MNAs) impose limit on the usability of the photoplethysmogram (PPG), particularly in the context of ambulatory monitoring. MNAs can distort PPG, causing erroneous estimation of physiological parameters such as heart rate (HR) and arterial oxygen saturation (SpO₂). In this study we present a novel approach “TifMA” based on Time-frequency spectrum of PPG to first detect the MNA corrupted data and next discard the non-usable part of corrupted data. The term “non-usable” refers to those type of PPG data of which HR signal cannot be recovered accurately. Two sequential classification procedures were included in TifMA algorithm. The first classifier detect between MNA corrupted and MNA free PPG data. Once a segment of data belong to MNA corrupted category, a new classifier is adopted to determine whether the HR can be recovered from the corrupted segment or not. A support vector machine (SVM) classifier was utilized to build a decision boundary for first classification task using data segments from a training data set. Features from time-frequency spectrum of PPG were extracted to build the detection model. Five database were considered for evaluating of TifMA performance: (1) and (2) Lab controlled PPG recordings from forehead and finger pulse oximeter sensors with random movements, (3) and (4) Real patients PPG recordings from UMass Hospital with random free movements and (5) Lab controlled PPG recordings from forehead while running on a treadmill. The first dataset was to analyze the noise sensitivity of the algorithm. Databases 2-4 were used to evaluate the MNA detection phase of algorithm. The results of first phase of the algorithm (MNA detection) were compared to three existing MNA detection algorithms: the Hjorth, kurtosis-Shannon Entropy and time-domain variability-SVM approach, an approach recently developed in our lab. The proposed TifMA algorithm our method consistently provided higher detection rates 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 an error of less than 1 sec in duration, whereas the next best algorithm had a detection error of more than 2.2 seconds. A final more challenging dataset were collected to verify the performance of the algorithm in discriminating the corrupted data that are usable for accurate HR estimations and those that are non-usable.

Index Terms—Time-Frequency, Motion and Noise Artifacts, Photoplethysmography, Complex Demodulation, Heart Rate Estimation

I. INTRODUCTION

PULSE oximeter (PO) is a non-invasive, low cost device that is widely used in hospitals and clinics to monitor heart rate (HR) and arterial oxygen saturation (SpO₂). Recently, there have been efforts to derive other physiological parameters from Photoplethysmogram (PPG), as recorded by a PO [1]–[3]. The fluctuations observed in a PPG are influenced by arterial, and venous blood, as well as autonomic and respiratory systems of the peripheral circulation. Such information could be used to more comprehensively phenotype cardiovascular health. Due to increasing health care costs, a single sensor from which multiple clinical data-points can be derived such as a 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 signs. Knowledge of respiratory rate and HR patterns would provide useful clinical information in many situations where a PO is the sole available monitor. However, extraction of the above mentioned vital signs and other physiological parameters using PO is predicated on artifact-free PPG data. It is well known that the PPG is highly sensitive to artifacts, particularly those generated while the patient is in motion [4]. This imposes a huge limitation on the usability of the PPG for ambulatory monitoring applications. Motion and noise artifacts (MNAs) distorting PPG recordings can cause erroneous estimation of HR and SpO₂ [5]. Although 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 MNAs in PPG has been the core focus of research for many years.

Although there are techniques which have been proposed to alleviate the effects of MNAs, 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 [6]–[12]. 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 and some algorithms are solely designed to capture only the HR and SpO₂ information instead of the signal’s morphology and its amplitudes, which are needed for other physiological derivations [13]. Moreover, these reconstruction algorithms operate even on clean PPG portions where MNA reduction is not needed. This introduces unnecessary computation burdens 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 [14].

MNA detection methods are mostly based on a signal quality index (SQI) which quantifies the severity of the artifacts. Some approaches quantify the SQI using waveform morphologies [15]–[17] or filtered output [18], [19], while other derive the SQI with the help of additional hardware such as accelerometer and electrocardiogram [20], [21]. In some commercially available off-the-self pulse oximeters, accelerometers are not readily present, or even if they are available, access to the raw data is not usually feasible, hence, they cannot be utilized for MNA cancellation. Moreover, traditional approaches to cancellation of MNA using adaptive noise filtering do not always yield accurate results.

Statistical measures, such as skewness, kurtosis (K), and quadratic phase coupling[22], Shannon entropy (SE), and Renyi’s entropy [23], have been shown to be helpful in determining the SQI. These statistical algorithms discriminate amplitude distributions between PPG segments with an assumption that clean and corrupt segments would form two separate groups. However, PPG morphology vary among patients, thus yielding multitude of amplitude distributions. Therefore, it would be difficult to obtain high accuracy from these algorithms in practice. Another approach using Hjorth parameters, where H1 and H2 represented central frequency and half of the bandwidth of a signal respectively, were proposed as means to quantify the degree of oscillation in a signal [24], [25]. They were employed for MNA discrimination in PPG by Gil et al. [26] with hypothesis that MNA derived H1 and H2 would largely differ from that of original PPG. However, due to time-varying dynamic nature of PPG, frequency features such as H1 and H2 alone would not be sufficient for accurate detection of MNA. Our recently published MNA detection method uses time-domain features such as variability in heart rate, amplitude, and waveform morphology with the help of the support vector machine (SVM) classifier for detection [27]. The algorithm, which we termed time-domain variability SVM (TDV) 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 the electrocardiogram (ECG), the 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 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 [28]–[30]. Yan et al. used smoothed pseudo Wigner-Ville TF technique for MNA reduction albeit with limited success [10]. In this paper, we introduced a novel algorithm for MNA detection utilizing TF representation produced by VFCDM. It is hypothesized in the design of our proposed MNA detection algorithm that TF information would provide meaningful dynamic features for improved differentiation of MNAs.

In this paper, we present a new MNA detection algorithm “TifMA” which not only has the potential to detect MNA corrupted PPG segment but also is able to discriminate between usable versus non-usable PPG segments. TifMA algorithm was developed based on the features from Time-frequency spectrum of PPG signal. Variable frequency demodulation technique was used to derive the time-frequency spectrum of PPG. The proposed algorithm includes two phases: (1) MNA detection, (2) Usability of the MNA corrupted data segment detection. The algorithm performance was evaluated at each phase using different PPG recordings. We show that features from time-frequency spectrum of PPG have a great potential in discriminating between MNA corrupted and clean PPG data while on the other hand it can provide a new ability to determine if the corrupted data are usable or not for estimating HR. Typically features that are extracted from VFCDM TF were used as inputs for the machine-learning classifier algorithm which utilized the Support Vector Machine (SVM). The results of MNA detection phase of TifMA were compared to three other existing MNA detection algorithms: the Hjorth [26], kurtosis-Shannon Entropy [23] and time-domain variability-SVM approach, an approach recently developed in our lab [27]. The output of usability detection stage of the algorithm obtained according to a fixed threshold value and were evaluated by comparing the reference HR to the estimated HR from VFCDM time-frequency spectrum.

II. MATERIALS AND METHOD

A. Experimental Protocol and Preprocessing

To develop, analyze and evaluate the proposed TifMA algorithm we used 5 different datasets. The datasets (2) and (3) were recorded in controlled condition from 10 subjects in Chon lab and the datasets (4) and (5) were provided by UMASS

Hospital recorded from 10 patients. Dataset (5) was also recorded in Chon Lab during a treadmill experiment. The details of experiments and data collection protocols are as follow

(1) and (2) [Chon Lab Dataset]: 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). Laboratory data allows us to have more control over the duration of MNA generated to ensure that the detection phase of algorithm was tested on a wide range of MNA duration. 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. The subjects were all healthy with no past histories of cardiovascular diseases. 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, a 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.

(3) and (4) [UMASS Hospital Dataset]: The next data set was acquired from patients who were admitted to our partner hospital at the UMass Memorial Medical Center (UMMC). 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. The patient PPG data were recorded from 10 subjects admitted to emergency 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. One subject had hypertension and 2 subjects suffered from hypercholesterolemia and hyperlipidemia; the remaining subjects were considered free from cardiovascular diseases. All recording were performed in the Emergency Department with a room temperature of 68 °F. 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.

(5) Chon Lab Dataset: This dataset was also recorded in the Chon Lab from 10 healthy subjects (9 male/1 female), with ages ranging from 26 to 55. For each subject, the PPG signal was recorded from their forehead using a PO. The ECG signal was recorded as a reference from the chest using ECG sensors, sampled at 400Hz. During data recording, subjects walked, jogged and ran on a treadmill with speeds of 3, 5 and 7 mph, respectively, for 9 min. At the end, all experimental subjects were asked to perform random arbitrary movements for 1 min. Activities involved are 1 min rest, 1 min walking (3 mph), 1 min rest, 2 min jogging (5 mph), 1 min rest, 2 min running (7 mph), 1 min rest, 1 min arbitrary movement. The ECG-based reference HR was recorded in order to assess the performance of the second phase of TifMA algorithm.

This study was approved by both WPI's and UMMC's IRBs and all subjects were given informed consents prior to data recordings. Chon Lab data used in this paper were collected using our custom-made multi-channel pulse oximeters. The subjects in the lab were all healthy with no past histories of cardiovascular diseases. All lab recordings were performed in a quiet room with temperature of 70 °F. There were two versions with different form factors to capture PPG at the user's forehead and finger. The forehead sensor is termed 6PD-forehead since it consisted of 6 photodetectors concentric around the center-paired LEDs with the peak wavelengths of 660nm (red) and 940nm (infrared). The finger sensor, due to space constrain, consisted of 3 photodetectors concentric around center paired LEDs with the same specification of the 6PD-forehead sensor. The PPGs were sampled at 80Hz. The scope of this study was to evaluate the efficacy of our proposed TifMA algorithm. Therefore, only one infrared PPG channel deemed to contain the best signal quality data was used for analysis.

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. Further details of this study's database including demographic of subjects are given in Table (1).

TABLE 1. PPG DATASETS AND EXPERIMENTS SETTINGS

No. Recordings	Dataset	Electrode Type	Subject's Age/Sex	Ethnicity
11	1 (Chon Lab)	Forehead Pulse Oximeter	23-58 y (9 Male, 1 Female)	3 Asian 2 Hispanic 3 American 2 White
11	2 (Chon Lab)	Finger Pulse Oximeter	23-58 y (9 Male, 1 Female)	3 Asian 2 Hispanic 3 American

				2 White
10	3 (UMass Hospital database)	Forehead Pulse Oximeter	18-38 y (5 Female, 5 Male)	3 African-American (3 female) 6 Caucasian (1 females/5 male) 1 Puerto Rican (female)
10	4 (UMass Hospital database)	Finger Pulse Oximeter	18-45 y (4 Female, 6 Male)	1 African-American 1 Hispanic 7 Caucasian 1 Puerto Rican
10	5 (Chon Lab)	Forehead Pulse Oximeter	23-58 y (9 Male, 1 Female)	3 Asian 2 Hispanic 3 American 2 White

A. Reference signal for MNA Detection

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 [22], [23], [27]. 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. Disagreements of the marked portions were resolved by majority votes. Cohen's κ was used to determine if there was agreement between three inspectors' judgement on whether PPG segments were declared to be clean or corrupted. For each dataset, the average Cohan's κ was reported in Table (2) over three runs computed from each distinct inspector pair. Overall, the Cohan's κ showed substantial agreement between the inspectors with 95% CI and $p < 0.0005$.

TABLE II. AVERAGED COHEN'S κ COEFFICIENTS REPRESENTING THE AGREEMENT BETWEEN INSPECTORS' DECISION ON PPG LABELS.

	Lab. Head	Lab. Finger	Umass. Head	Umass. Finger
Cohan's κ	0.823	0.935	0.732	0.791
# of subjects	11	11	10	10
Duration (min)	5	5	10	10

Due to the differences with operating window lengths between the comparing methods, detection results and reference labels were converted to every second. For example, a 1 minute PPG segment would have 60 detection values from each method and 60 reference labels. These results were then used to compute sensitivity, specificity, accuracy for each method.

B. Reference signal for Data Usability Detection

A usability index was calculated to measure the usability of MNA corrupted segments of PPG recordings. In order to verify the performance of the index, it was compared to the reference usability measurements. The reference was determined according to the deviation of reference HR from chest, and the estimated HR obtained from phase 2 of TifMA algorithm.

C. TifMA

As mentioned above TifMA algorithm consists of two major phases: (1) MNA detection, (2) Usability detection. Both phases of algorithm were developed based on a time-frequency technique named VFCDM. 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 [31]. VFCDM has two steps: (1) constructing an initial TFS (iTFS) using a method developed in our laboratory, termed fixed frequency complex demodulation (FFCDM); (2) the center frequencies of the iTFS are used for further complex demodulation (CDM) to obtain even more accurate TFS and amplitude of the TFS. The VFCDM methodology is presented in Table (3).

TABLE III. VFCDM ALGORITHM PROCEDURE

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 $dc(t)$:

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

Step (1) For a given center frequency, we can extract the instantaneous amplitude information $A(t)$ and phase information $\phi(t)$ by multiplying (1) by $e^{-j \int_0^t 2\pi f(\tau) d\tau}$ which results in the following:

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

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

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

Step (3) By changing the center frequency followed by using the variable frequency approach as well as the LPF, the signal, $x(t)$, will be decomposed into the sinusoid modulations, d_i , by the CDM technique as follows:

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

Step (4) The instantaneous frequency and amplitude of d_i can be calculated using the Hilbert transform

$$A(t) = 2|z_{lp}(t)| = [X^2(t) + Y^2(t)]^{1/2} \quad (5A)$$

$$X(t) = \text{real}(z_{lp}(t))$$

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

$$\phi(t) = \arctan\left(\frac{\text{imag}(z_{lp}(t))}{\text{real}(z_{lp}(t))}\right) = \arctan\left(\frac{Y(t)}{X(t)}\right) \quad (5B)$$

$$f(t) = f_0 + \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (5C)$$

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.

1) TifMA Phase (1): MNA Detection

The first phase of algorithm time-frequency spectrum (TFS) of PPG is utilized to extract features that contribute in discriminating between MNA corrupted PPG data and clean data segments. Fig. 1 (A) and (B) illustrates an example of VFCDM-TFS of 4 sec segment of clean and MNA corrupted PPG data. It can be observed from this figure that TFS of a corrupted PPG segment has different characteristics comparing to TFS of the clean segment. The heart rate trace and its harmonics are distinguishable in the clean TFS while motion and noise artifacts distort the HR and its harmonics' traces in TFS of MNA corrupted PPG. In this paper FM_1, FM_2 and FM_3 terms are referred to the HR trace and two of its harmonic traces, while to refer the corresponding spectral power, the terms AM_1, AM_2 and AM_3 are used. FM, and AM stands for frequency modulation and amplitude modulation respectively. Since respiratory induced fluctuation in PPG is highly dynamic and not trivial to characterize in a TFS, regions associated with respiratory frequencies are removed from the TFS by setting their powers to zero. The respiratory frequencies are defined to be between 0 – 0.5Hz. Our algorithm first determines the dominant frequency in the PPG segment termed f_1 as predicated that PPG is dominantly driven by cardiac cycles. To determine instantaneous f_1 , total powers within narrow band spectral window $W_k = [f_k - BW, f_k + BW]$ are computed: where $BW = 0.2 \text{ Hz}$ is the bandwidth of the band; f_k is the center frequency of the window ranging from 0.6Hz – 2.4Hz with increment of 0.1Hz. BW was determined based on an assumption that within a short time HR does not fluctuate more than 12 beat per minute. f_1 is estimated as the f_k at which W_k is maximum.

The TFS of the segment is normalized by the total power in the f_1 band. It then extracts FM_1 and AM_1 within $[f_1 - BW, f_1 + BW]$. Note that each point in the TFS has three instantaneous values: time, frequency, and spectral power. The maximal power in each time instance is taken to form $AM_1 = AM_1(t): t = 1, \dots, N - 2$ where N is a number of data-points in the PPG segment. Its corresponding FM_1 is also extracted. Once located, AM_1 is removed from the TFS by setting its power to zero. Similarly, $\{AM_2, FM_2\} \in [2f_1 - BW, 2f_1 + BW]$ and $\{AM_3, FM_3\} \in [3f_1 - BW, 3f_1 + BW]$ are found and removed. Note that, the proposed algorithm assumes that corrupted PPG would exhibit irregularity in the time series waveform. Thus, the TFS of a corrupted segment would have broadband dynamics. In this case, the proposed fundamental frequency estimation method would probably yield inaccurate result. However, an arbitrary estimation of HR (fundamental frequency and its harmonics) in such cases would remove only a portion of the noise power within the defined bands, but would retain the rest.

From the extracted TFS and $\{FM_i, AM_i: i = 1, 2, 3\}$, three time-frequency (TF) features were derived to quantify the noise level between clean versus corrupted PPG segments.

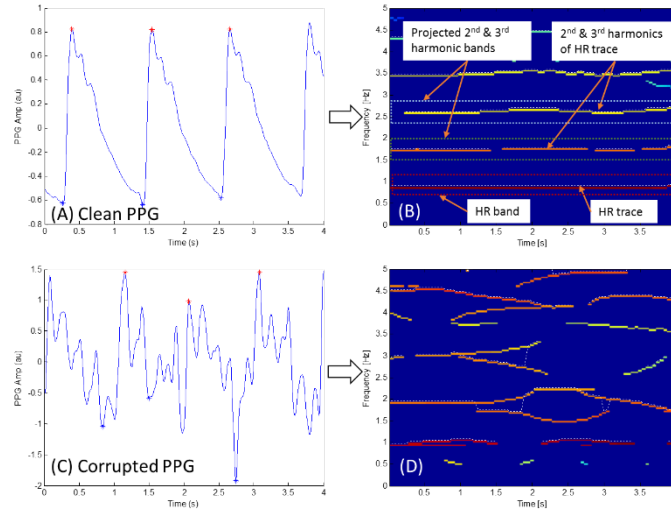


Figure 1. Time-Frequency Spectrum produced by VFCDM in 8sec PPG window ($L = 4s$). The rectangular boxes are fundamental band and its 2nd and 3rd projected harmonic bands. The white dotted lines inside their corresponding bands are extracted fundamental and harmonic traces of the PPG segment. (A) Clean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

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:

$$P_{noise} = P_{TFS} - \sum_{i=1}^3 \sum_t AM_{i,t} \quad (6)$$

where P_{noise} is the total power in the TFS. In a clean PPG segment as illustrated in Fig. 1B, 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 as illustrated in Fig. 1D. Some of these spectral power are outside of the bands and/or there are multiple powers within a band. Therefore, these powers 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:

$$df_{FM} = \sum_{i=2}^3 \sum_t |FM_{i,t} - i \times FM_{1,t}| \quad (7)$$

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. This feature measures the agreement between the fundamental frequencies detected from the TFS and from the time domain signal. It is assumed that the frequencies would be in agreement in a clean PPG segment. But there would be potentially large difference in corrupted segment. df_{HR} is computed as:

$$df_{HR} = \sum_t \left| FM_{1,t} - \text{median} \left(\frac{1}{PP} \right) \right| \quad (8)$$

where PP (sec) are the peak-to-peak intervals in a PPG segment.

As mentioned above the results of MNA detection phase of TifMA were compared to three other existing MNA detection algorithms: the Hjorth [26], kurtosis-Shannon Entropy [23] and time-domain variability-SVM approach, an approach recently developed in our lab [27]. Table (4) present the features these methods extract from PPG data.

TABLE IV. FREQUENTLY USED MNA DETECTION METHODS IN LITERATURE

- **Kurtosis and Shannon Entropy detector (KSE)**

KSE algorithms utilized two statistical measures, Kurtosis and Shannon Entropy, that quantify distributions of PPG segment [23]. Kurtosis (K) describes the distribution of observed data around the mean. K is defined as:

$$K = \frac{E(x-\mu)^4}{\sigma^4} \quad (10)$$

where μ is the mean of x , σ is the standard deviation of x , and $E(t)$ represents the expected value of the quantity t . Shannon Entropy (SE) quantifies how much the probability density function of the signal is different from a uniform distribution. SE is defined as:

$$SE = - \sum_{i=1}^k \frac{p(i) \log(p(i))}{\log(\frac{1}{k})} \quad (11)$$

Where i represents the bin number; $k = 16$; and $p(i)$ is the probability distribution of the signal

- **Hjorth detector**

Gil et al. used Hjorth parameters with an assumption that when the signal differs largely from an oscillatory signal, it is very likely an artifact [26]. Hjorth parameters are defined from the i^{th} -ordered spectral moment $\bar{w}_i$:

$$\bar{w}_i(n) = \int_{-\pi}^{\pi} w^i S_x(e^{j\omega}) d\omega$$

$$H_1 = \sqrt{\frac{\bar{w}_2(n)}{\bar{w}_0(n)}} \text{ and } H_2 = \sqrt{\frac{\bar{w}_4(n)}{\bar{w}_2(n)}} \quad (12)$$

where $S_x(e^{j\omega})$ is the power spectrum of a PPG segment $x(n)$. H_1 and H_2 represent the central frequency and half of bandwidth respectively. amplitude.

- **Time-domain variability and support vector machine (TDV)**

TDV algorithm utilized four time-domain features to describe the variability in clean versus MNA corrupted PPG segments [27]. The features are:

- Variance in HRs (TDV_{HR})
- Variance in pulse amplitudes (TDV_{AMP})
- Variance in systolic-diastolic ratios (TDV_{SD})
- Variance in pulse shapes (TDV_{WAV})

Feature extraction from PPG was done on a sliding window segment of length $L_1 = 8$ seconds with 50% overlap. Each segment was transformed using VFCDM into TFR, of which only the middle portion of length $L_2 = 4$ seconds was considered for further processing. This is because as shown in the Results section, the middle 4 second data length from the initial $L_1 = 8$ seconds for the subsequent VFCDM analysis provided the best accuracy in detection of MNAs.

To accurately pinpoint the time occurrence of MNAs, we implemented a *trace-back strategy*, which is triggered when the P_{noise} value changes its state as illustrated in Fig 2. When P_{noise} goes from lower than a threshold value of 0.15 to greater than 0.15, the trace-back algorithm updates new TF features three times with shifting backward a second at each time instant. For example in Fig. 2A, P_{noise} changes to a value that is greater than 0.15 at time duration 4-8 seconds. The trace-back scheme would call the VFCDM routine to compute new TF features for the back-shifted segments at time durations starting at 3-7 seconds, 2-6 seconds, and ending at 1-5 seconds. 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 MNAs. 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. An example of the track-back strategy on an actual PPG signal is illustrated in Fig. 2 where: (A) shows TF features are updated at a possible starting point of a MNA-corrupted segment; and (B) shows TF features are updated at a possible ending point of a MNA-

corrupted segment.

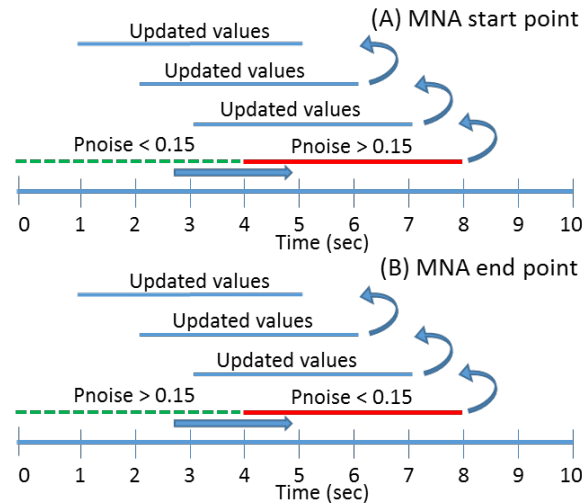


Figure 2. Trace-back strategy to find the start (A) and end (B) points of MNA.

Support Vector Machine (SVM) was applied to build a decision boundary to detection MNA corrupted PPG segments from clean data. SVM is widely used for classification and regression analysis due to its accuracy and robustness to noise [12], [32].

Fig.3 shows a representative illustration of the performance of our proposed approach. Fig.3A displays a pre-processed PPG containing both clean and MNA-corrupted periods. Fig.3B-D depict the corresponding *TF features* explained in Section B.2. As shown, the *TF features* have low values for the clean portion of the data whereas it is high where MNAs occur. The *TF features* are used for the SVM classifier to determine whether the given segment is clean or corrupted. Fig.3F shows the classification results.

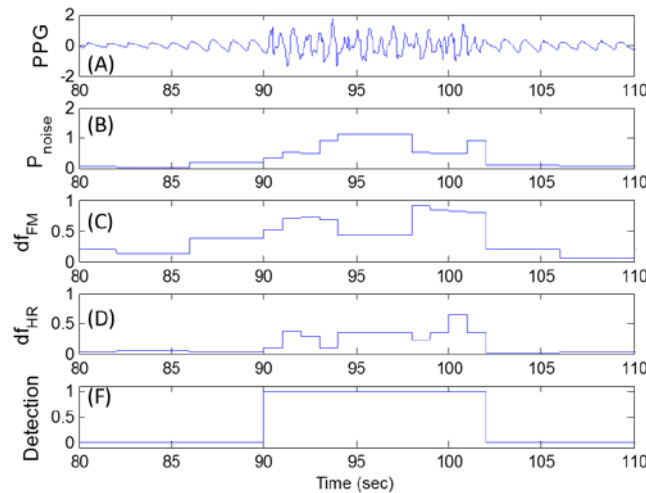


Figure 3. An example of MNA detection using VFCDM. (A) PPG signal corrupted with MNA. (B) *noiseQI* derived from VFCDM. (C) Detection decision of VFCDM (blue) and reference MNA (red).

A representative example of the MNA detection comparing all of the aforementioned methods is illustrated in Fig. 4.

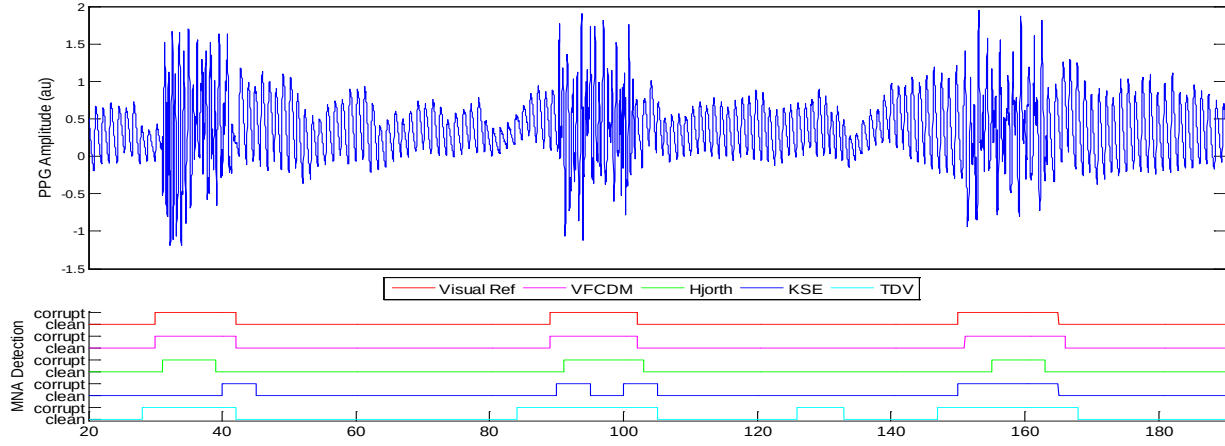


Figure 4. An example of MNA detection using our VFCDM method versus other methods: Hjorth, Time-domain and Kurtosis-Shannon Entropy. The pulse-like traces are the MNA reference and detection results from the feature sets. High value indicates detected MNA, otherwise clean PPG signal.

2) TifMA Phase (2): Usability Detection

From the previous section we show that when the PPG signal is corrupted, the HR trace and its harmonics in TFS also exhibit distortion but the amount of distortion is not uniform. Thus, a proper way to determine if a corrupted signal is usable for accurate HR estimation or not is to extract the tracing of HR values from TFS and analyze the tracing time series for missing values and abrupt changes in HR estimations. In order to fully investigate and evaluate the above described idea, a more challenging dataset with slow and rapidly changing HR scenarios was adopted in this phase of the study. The data were recorded from subjects while walking, jogging and running on a treadmill. Fig.5 illustrates a 7.5 min (0.5^{min} rest, 2^{min} walking (3mph), 1^{min} rest, 2^{min} jogging (5 mph), 1^{min} rest, 1^{min} running (7 mph)) segment of the 9 min recordings from subject 1. It can be observed from the reference HR (Fig.5D) estimated from a five-lead Holter monitor with hydrogel ECG electrodes (Fig.5A) that the HR is changing during this episode of experiment in a range of 80-160 bpm. Fig.6 represents the VFCDM-TFS of the corresponding PPG data in Fig.5B. One can identify the tracing of HR in TFS which begins to show distortion during the timestamps of 260-340 sec and 440-500 sec. By applying the MNA detection procedure of TifMA one can see that motion related segments truly marked as MA corrupted segments. Now we would like to investigate if any part of this MNA corrupted segments is still usable for accurate HR estimation.

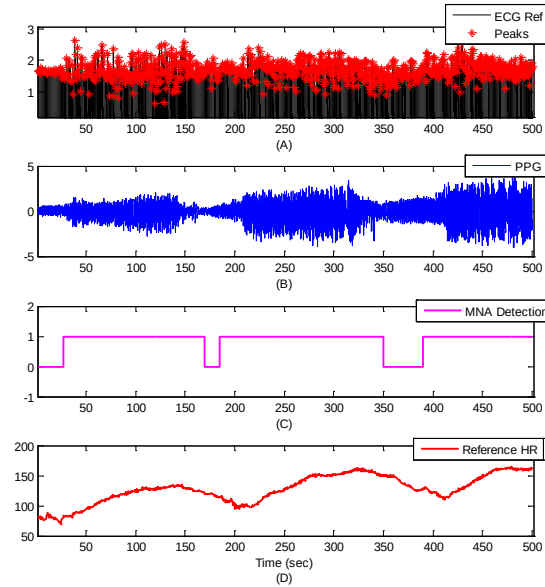


Figure 5. Subject # 7 from dataset (3): (A) reference ECG, (B) PPG recordings, (C) TifMA detection, (D) Estimated reference HR from reference ECG

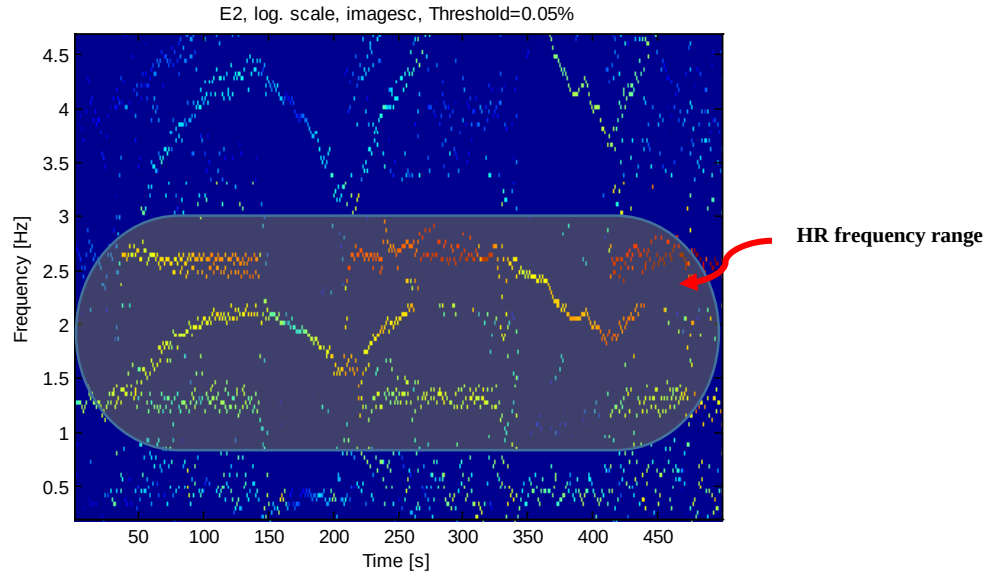
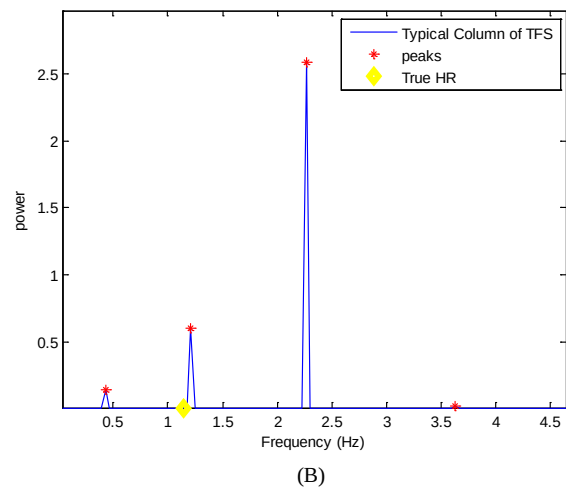
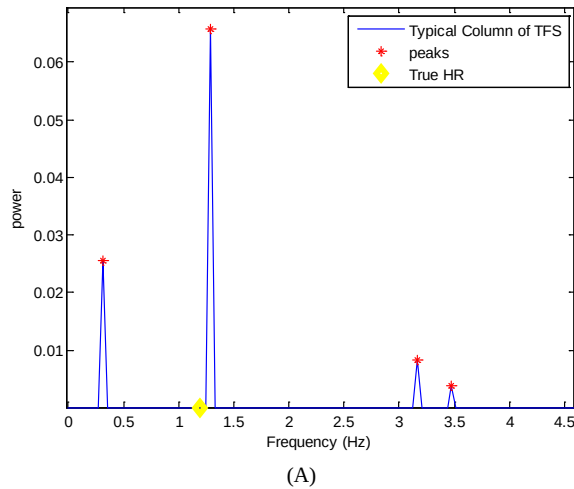
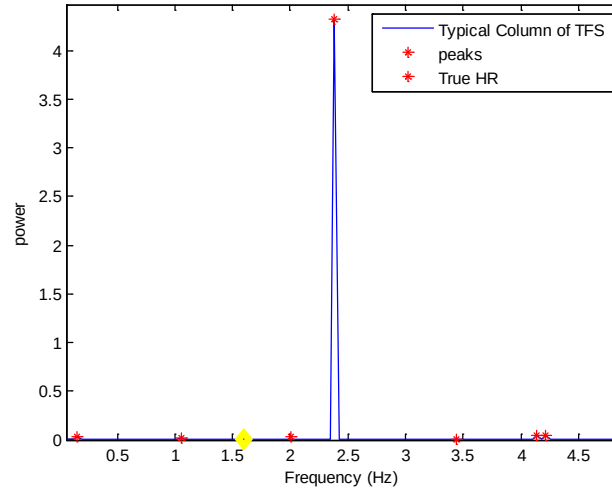


Figure 6. VFCDM Time-Frequency Spectrum of Subject #7 from dataset (3)

The following describes the second phase of TifMA algorithm, which is detection of non-usable and usable parts of the identified corrupted data segments.

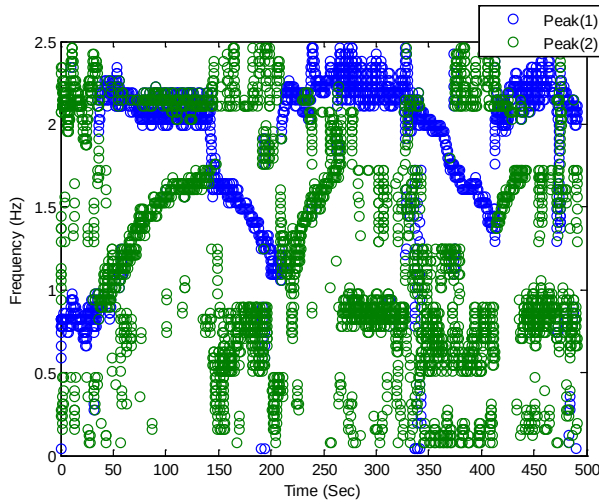
The usability detection stage is comprised of three steps: (1) TFS filtering, (2) HR tracking and extraction (3) Usability index measurement. In TFS filtering step, the TFS spectrum is reduced to a two component TF. Fig.7 is the plot of three typical columns of PPG TFS matrix in Fig.6. In the Fig.7A and B, the true HR frequency is close to the first and second peak respectively, while in the Fig.7C the true HR frequency is far from the dominant peaks in the spectrum. Thus, the first two those type of data that is usable for HR estimation but the third one is not usable at all. We can assume that as long as the PPG data is clean the HR frequency belong to the frequency component with highest power (peak) in the each column of TFS matrix. On the other hand when the data get corrupted by MNA, it might shift the HR frequency to the next highest peak in the spectrum or might lose HR frequency. Here we proposed to design a TFS filtering step to look into the first two highest peaks power and frequencies in each column of TFS. Hence, the original TFS (Fig.6) can be filtered to keep only the prominent components of spectrum (see Fig.8A).



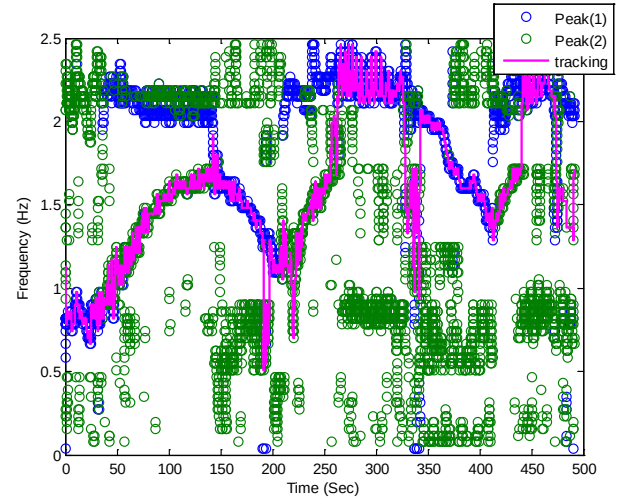


(C)

Figure 7. Example of usable and non-usable PPG data: (A) Usable: True HR frequency is close to the first highest peak of spectrum, (B) Usable: True HR frequency is close to the second highest peak, (C) Non-usable: True HR is not close to any of prominent peaks in the spectrum



(A)



(B)

Figure 8. Filtered TFS: (A) First and second prominent components of TFS shown in Fig.5, (B) HR Extraction from filtered TFS

The next step after filtering the TFS is extracting HR from it. The HR tracking procedure is straightforward as follow: Assuming that we have the knowledge of the initial HR, the HR at each window is extracted by comparing the peaks to the previous HR value, if either of the peaks are in 0.8 Hz (5 bpm) range from the most recent value of HR it is chosen as usable if the value deviates more than (10 bpm) it is considered as non-usable. Fig.9A represents the estimated moving averaged HR from TFS comparing to the reference HR from ECG. The reference of usability detection procedure is set according to the deviation of TifMA estimated HR and the true reference HR (see Fig.9B). Fig.9C shows the results of MNA detection, the first phase of the algorithm. It can be observed from this figure that segments of PPG recordings during movement (walking/jogging/running) has been mostly detected as noisy. Fig.9D shows the usability index derived by second phase of TifMA. One can observe by comparing the reference UI and TifMA's UI, that TifMA is detecting the usable/nonusable portions of data for HR estimations very well. The accuracy, specificity and sensitivity of the proposed algorithm on all of the 10 recordings from dataset#5 is presented in the following section.

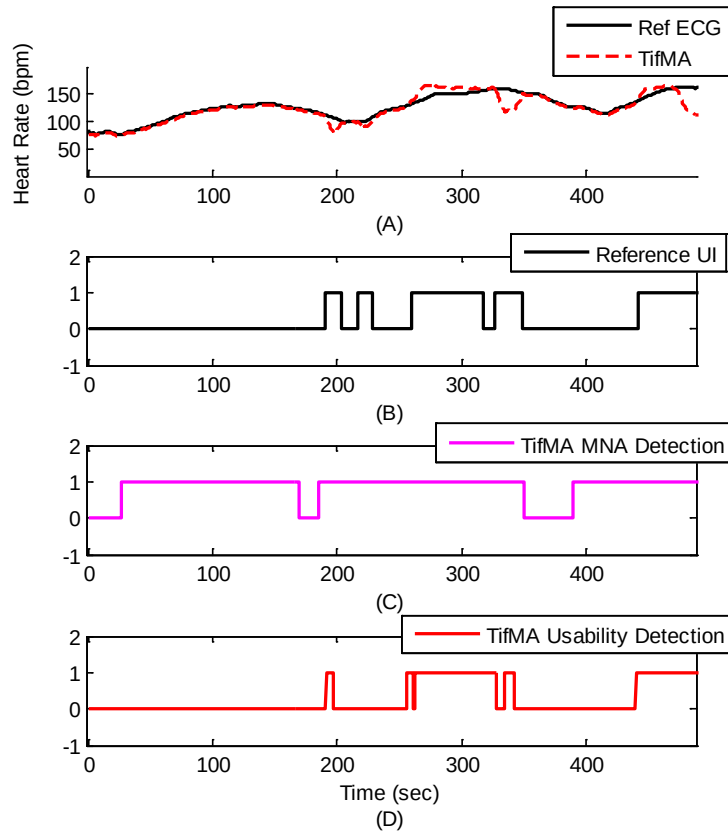


Figure 9. Reference usability index. (A) Comparison of TifMA estimated HR to reference HR from ECG, (B) Reference Usability Index (UI) as an indicator of HR estimability from TFS during motion, (C) TifMA MNA Detection Result, (D) TifMA Usability Detection

The overall TifMA algorithm flowchart is depicted in Fig.10.

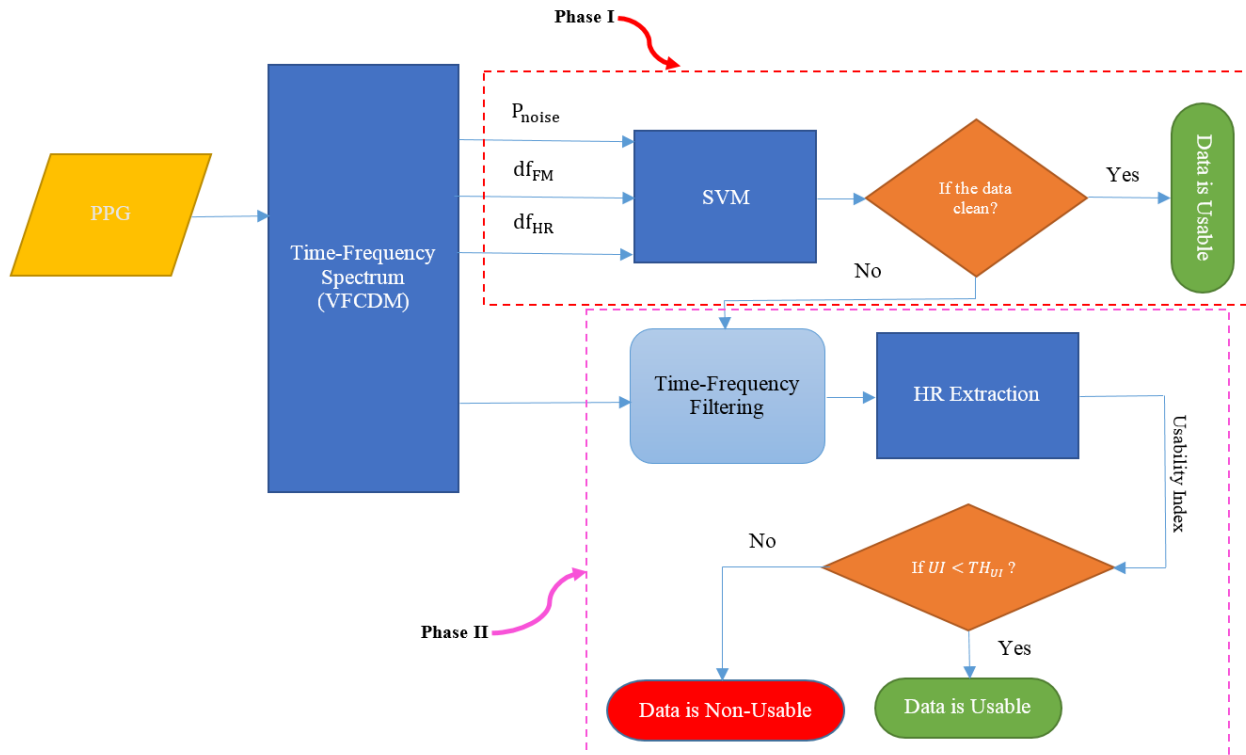


Figure 10. TifMA Algorithm Flowchart

III. RESULTS

We evaluated the performance of the TifMA algorithm on 5 datasets. The first 4 datasets (laboratory controlled, and hospital patients with motion-corrupted PPGs) were used to measure the performance of the first phase of the algorithm (MNA detection) and the fifth (treadmill experiment) dataset was used to evaluate the second phase of algorithm (usability detection).

A. MNA Detection Results

Leave-one-out cross validation was adopted to evaluate the performance of MNA detection phase [33]. Specifically, for a dataset of N subjects, data from $N-1$ subjects were used for training and the remaining subject data were used for testing. The train-test cycle was repeated 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.

The optimal window length was determined by varying L_2 from 3 to 6 sec while keeping L_1 constant at 8 sec, as described in phase one of the TifMA. Detection performance was evaluated by comparing our classification results to the MNA reference (as determined visually by the experts) to yield accuracy, sensitivity, and specificity. Table (5) shows performance statistics in terms of accuracy (Acc), sensitivity (Sen), and specificity (Spe) of our MNA detection algorithm at various window lengths (L_2) for the laboratory collected dataset (1). The window length of 4 sec ($L_2 = 4$ second) yielded the best performance in term of accuracy among the various window lengths.

TABLE V. MEAN $\pm$ STD. DEVIATION OF PERFORMANCE METRICS OF OUR PROPOSED TIFMA USING VARIOUS WINDOW LENGTH.

L		3s	4s	5s	6s
Lab. Head	Acc	86.8 $\pm$ 5.52	95.7 $\pm$ 0.82	86.9 $\pm$ 4.27	85.5 $\pm$ 6.96
	Sen	82.0 $\pm$ 9.59	93.0 $\pm$ 5.75	81.5 $\pm$ 8.47	79.8 $\pm$ 9.58
	Spe	89.1 $\pm$ 5.86	96.6 $\pm$ 1.48	81.5 $\pm$ 3.19	90.0 $\pm$ 5.97

We compared the proposed algorithm with three other published MNA detection algorithms: 1) Hjorth features (Hjorth); 2) time-domain variability SVM (TDV) approach; 3) Kurtosis-Shannon Entropy (KSE) features [22], [23], [26], [27]. Table (6) represents the performance results of each method in term of means and standard variations of the accuracy, sensitivity, and specificity results from the cross-validation. Kruskal-Wallis H tests were done to determine if there was a statistical difference in accuracy, sensitivity, and specificity among different detection methods. If there was a statistical difference, Mann-Whitney post-hoc test was performed between our proposed TifMA method and each of the compared methods.

In order to measure and compare the detection powers, receiver operative characteristic (ROC) curves were generated for all the features used in TifMA and other detection algorithms. Area under these curves (AUCs) represent the strength of these features. In Fig. 11, ROC curves of the three TF features as well as the AUCs of all features used for MNA detection are shown.

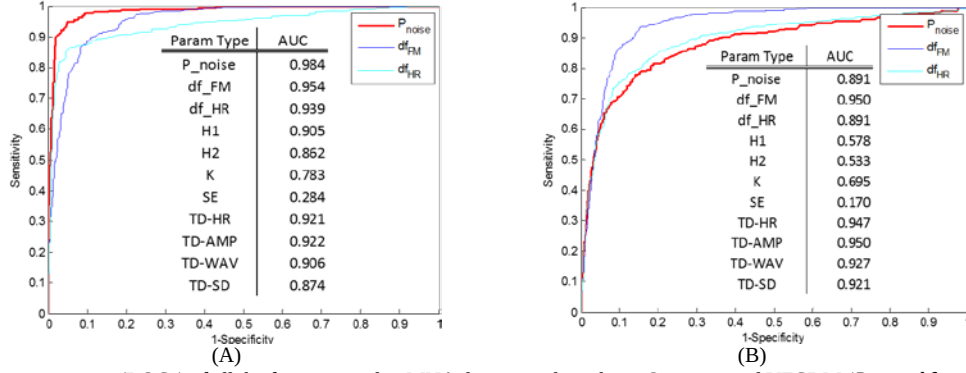
TABLE VI. MEAN $\pm$ STD. DEVIATION OF PERFORMANCE METRICS OF OUR PROPOSED TIFMA, OTHER METHODS. (*) INDICATE STATISTICAL SIGNIFICANCE ($p < 0.05$) BETWEEN OUR METHOD VERSUS THE OTHERS.

		TifMA	Hjorth	TDV	KSE
Lab. Head	Acc	95.7 $\pm$ 0.82	72.5 $\pm$ 10.74 *	93.4 $\pm$ 3.50 *	83.1 $\pm$ 7.31 *
	Sen	93.0 $\pm$ 5.75	47.2 $\pm$ 28.80 *	88.8 $\pm$ 7.90 *	56.6 $\pm$ 17.88 *
	Spe	96.6 $\pm$ 1.48	84.4 $\pm$ 4.63 *	96.7 $\pm$ 3.00	91.5 $\pm$ 1.14 *
Umass. Head	Acc	95.3 $\pm$ 1.34	69.5 $\pm$ 21.83 *	89.8 $\pm$ 1.45 *	85.1 $\pm$ 8.34 *
	Sen	90.8 $\pm$ 2.83	53.8 $\pm$ 26.05 *	84.6 $\pm$ 2.89 *	68.7 $\pm$ 17.72 *
	Spe	98.7 $\pm$ 1.07	84.5 $\pm$ 8.83 *	94.3 $\pm$ 3.95 *	86.4 $\pm$ 7.82 *
Lab. Finger	Acc	97.5 $\pm$ 1.50	91.1 $\pm$ 6.87 *	94.4 $\pm$ 3.30 *	58.5 $\pm$ 20.67 *
	Sen	96.4 $\pm$ 2.34	83.5 $\pm$ 21.04 *	94.7 $\pm$ 3.40	34.6 $\pm$ 12.23 *
	Spe	98.1 $\pm$ 1.43	96.2 $\pm$ 3.02	94.7 $\pm$ 3.00 *	86.3 $\pm$ 15.65 *
Umass. Finger	Acc	94.3 $\pm$ 1.64	71.0 $\pm$ 18.96 *	89.6 $\pm$ 2.34 *	88.3 $\pm$ 2.68 *
	Sen	88.5 $\pm$ 2.23	41.1 $\pm$ 27.57 *	85.2 $\pm$ 3.75 *	71.5 $\pm$ 8.83 *
	Spe	96.9 $\pm$ 1.86	88.4 $\pm$ 7.52 *	95.1 $\pm$ 0.79	93.6 $\pm$ 1.52 *

In addition to accurate MNA detection, the other attractive feature of our proposed algorithm is that it is able to accurately locate the start and end points of MNA occurrences. Accurate detection of the start and end time of the MNAs 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 the noisy portion of the PPG when the segment is designated to be clean. To evaluate the algorithm's effectiveness in pinpointing the start and end time of the MNAs, 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 accurately on average a MNA algorithm detects the start and end time of the MNAs. Table (7) provides DTT comparison of TifMA algorithm and other detection algorithms. As shown in Table (7), our algorithm's detection accuracy of the duration of the MNAs is significantly better than three other methods we compared. Our algorithm's DTT is less than one second whereas the second best algorithm, the Hjorth, is off by more than two seconds and the least accurate method, the KSE, is off by more than four seconds.

TABLE VII. MEAN $\pm$ STD. DETECTION OF TRANSITION TIME (DTT) OF TIFMA AND OTHER METHODS.

Algorithm	DTT (sec)
TifMA	0.91 ± 0.59
Hjorth	2.17 ± 0.37
KSE	4.24 ± 2.42
TDV	2.75 ± 0.96

Figure 11. Receiver-operating-curves (ROCs) of all the features used in MNA detection algorithms: Our proposed VFCDM (P_{noise} , df_{HR} , df_{FM}); Hjorth parameters (H1,H2); Statistical features (K,SE), Time-domain features (TD-HR, TD-AMP,TD-WAV,TD-SD). (A) Lab. Finger data; (B) Lab. Head data.

B. Usability Detection Results

By applying the first phase of TifMA algorithm we showed that the algorithm is able to determine if a PPG data segment is clean from MNA or is corrupted. However being corrupted by movement does not necessarily indicate that the whole noisy data is non-usable especially for HR monitoring applications. It was shown that HR frequency changes can be tracked using VFCDM time-frequency plot as long as the signal is usable and the level of corruption is not too high. This non-usable scenario can be occurred during any PPG recordings due to abrupt and fast movements that makes tracking of frequency harder in time-frequency domain. To this end the second phase of algorithm was tested on PPG recordings from 10 subjects who performed 9 min (walk/jog/run) experiment. The data initially was fed into the MNA detection phase and when a corrupted data segment is obtained, the second phase of algorithm perform usability measurement of data, that is whether the corrupted signal is usable for HR estimations or not. Table (8) presents the accuracy, sensitivity, and specificity of usability detection comparing to the reference usability index.

TABLE VIII. TIFMA USABILITY DETECTION PERFORMANCE.

Subjects	Dataset (5)	Acc	Sen	Spe	Non-usable period (min)
1		0.89	0.94	0.92	1.80
2		0.93	0.90	0.88	1.31
3		0.95	0.92	0.90	1.58
4		0.93	0.91	0.96	0.12
5		0.91	0.93	0.84	0.35
6		0.91	0.90	0.90	0.48
7		0.93	0.92	0.91	2.23
8		0.95	0.94	0.92	0.33
9		0.94	0.96	0.94	0.62
10		0.90	0.97	0.90	1.05
mean $\pm$ std		0.92 ± 0.02	0.93 ± 0.02	0.90 ± 0.03	0.98 ± 0.71

It can be seen from the above table that on average almost 10% of PPG recordings during a treadmill experiment were non-usable and should be discarded.

IV. DISCUSSION

We propose a novel MNA and usability detection method “TifMA” that uses dynamic characteristics of the corrupted PPG derived via the VFCDM. The algorithm comes in two phases: (1) MNA detection, (2) Usability detection. The efficacy of the detection phase was validated using contrived motion data from healthy subjects and unconstrained MNA data from participants recruited from a hospital-setting. The second phase of the algorithm was tested on HR varying scenario where subjects were asked to walk/jog/run for 9 min on a treadmill. For MNA detection several key features associated with MNAs derived from the VFCDM-based time-frequency spectrum. By transforming the PPG time series into the time-frequency domain, we were able to better capture time-varying characteristics of the MNAs. Specifically, we recognized that PPG’s clean signal dynamics are largely concentrated at the heart rate and its harmonic frequency bands. Hence, we surmised that the presence of large amplitudes in the other frequency bands must be associated with MNAs. This is clearly seen in Fig. 1B as VFCDM results from a clean PPG yields

distinct peaks across all times at the HR frequencies and its two successive harmonic frequencies. Therefore, we divided the TFS into three narrow band spectra and tracked down these frequency traces accordingly. In a clean PPG segment (shown in 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 values of *TF features*.

The detection accuracy on both lab-controlled and UMMC datasets using TifMA outperformed the other three detection methods: Hjorth, TDV-SVM, and KSE. We compared each method's performance based on their own unique feature selection by evaluating the area under the ROC curve. The AUCs showed that our *TF features* provided the highest values $AUCs > 0.89$ for both finger data and forehead recorded PPG, as shown in Fig. 5. Concomitantly, the accuracy, sensitivity and specificity values of our proposed method were significantly higher than other methods as indicated in Table III.

Since TifMA uses the features in frequency domain, not all noise dynamics in time domain are reflected in the frequency domain. Thus, time-domain motion noise artifact detection techniques [20, 21, 27] (e.g. accelerometer based MNA detection) are not always accurate when it comes to heart rate or respiratory rate monitoring applications. Fig. 12 illustrates an example of MNA analysis of PPG recordings from a subject during walking. Fig.12D shows the HR estimations from PPG and reference HR estimated from clean reference ECG. Fig.12C represents the accelerometer intensity of raw tri-axial accelerometer recordings. Movement intensity from three axial accelerometer signals is calculated by taking moving average of squared derivative of accelerometer raw data [21]. It can be observed from this figure that accelerometer is very sensitive to movement and as soon as the subject start to move (walk) it shows a level of increase in the accelerometer signal intensity. However by looking into Fig.12D, the HR from PPG can be accurately estimated in some section of movement period. By applying TifMA algorithm on this segment of PPG, the algorithm is able to discriminate between the usable and non-usable segments of PPG for HR estimations (see Fig.12B).

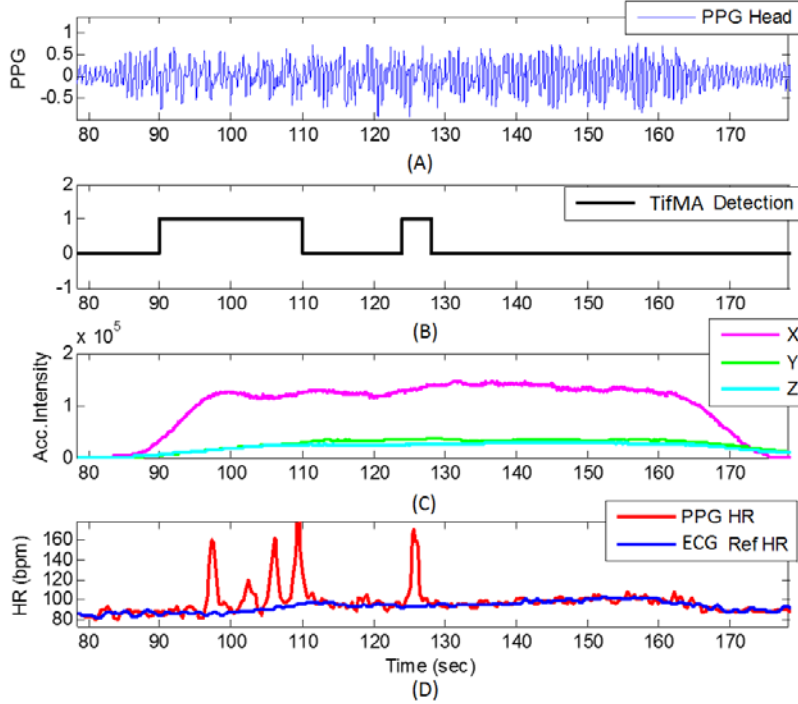


Figure 12. An example showing Accelerometer vs. TifMA based methods performance on a typical PPG signal recorded from a subject during walking.

It can be seen from the above example that accelerometers' high intensity values do not always lead to gross inaccurate HR estimation. For instance during the time window 95 to 115 sec of recordings, the HRs estimated from PPG significantly deviate from the reference ECG HR, as expected since the accelerometers intensity during this segment is quite high especially for the X-axis accelerometer. However, for a time segment from 130 to 160 sec, the HR estimation from PPG is very close to that of reference ECG even though the accelerometers' intensity is as high as the time window segment between 95-115 seconds. Therefore, just relying on accelerometers can lead to incorrect detection of MNA in the time segment between 130-160 seconds in Fig.12. Note however, with our proposed TifMA algorithm, it can pinpoint the start and end points of usable and non-usable segments of data that is it can correctly labeled the data segment between 130-160 seconds as usable data for HR accurate estimations.

The eventual aim of our proposed algorithm is to detect MNAs in real time. The algorithm only takes 33.3ms to compute *TF features* for a 4s PPG window length using Matlab running on a PC with the Intel Xeon processor operating at 3.6GHz. Therefore,

it would be straightforward to optimize the algorithm for real time detection of MNAs in PPGs.

In conclusion, we proposed an accurate MNA detection algorithm that utilizes both time and spectral features to classify between clean and corrupted PPG data segments. Moreover, it has the ability to go beyond MNA detection and can detect if the corrupted signal is still usable for accurate HR estimations or not. Comparison using four datasets showed our algorithm out-performed other contemporary MNA detection algorithms. Our algorithm also showed superiority with respect to detecting onset and offset of MNAs. Finally, TifMA is real-time realizable and it is applicable to either transmission (finger) or reflectance (forehead) recorded PPGs.

REFERENCES

- [1] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, "Automatic real time detection of atrial fibrillation," *Ann. Biomed. Eng.*, vol. 37, no. 9, pp. 1701–1709, Sep. 2009.
- [2] L. P. Antonsen and K. A. Kirkebøen, "Evaluation of fluid responsiveness: is photoplethysmography a noninvasive alternative?," *Anesthesiol. Res. Pract.*, vol. 2012, p. 617380, 2012.
- [3] F. Camacho, "Statistical Analysis of Central Aortic Blood Pressure Parameters Derived From the Peripheral Pulse," PhD Thesis, The University of New South Wales, Sydney, 2005.
- [4] 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, no. 6 Suppl, pp. S78–84, Dec. 2007.
- [5] R. Sahni, A. Gupta, K. Ohira-Kist, and T. Rosen, "Motion resistant pulse oximetry in neonates," *Arch. Dis. Child. Fetal Neonatal Ed.*, vol. 88, no. 6, pp. F505–F508, Nov. 2003.
- [6] 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," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1445–1457, May 2012.
- [7] T. L. Rusch, R. Sankar, and J. E. Scharf, "Signal processing methods for pulse oximetry," *Comput. Biol. Med.*, vol. 26, no. 2, pp. 143–159, Mar. 1996.
- [8] R. Hong Enríquez, M. Sautié Castellanos, J. Falcón Rodríguez, and J. L. Hernández Cáceres, "Analysis of the photoplethysmographic signal by means of the decomposition in principal components," *Physiol. Meas.*, vol. 23, no. 3, pp. N17–29, Aug. 2002.
- [9] S. M. A. Salehizadeh, Duy K. Dao, Jo Woon Chong, David McManus, Chad Darling, Yitzhak Mendelson, Ki H. Chon, "Photoplethysmograph signal reconstruction based on a novel hybrid motion artifact detection-reduction approach. Part II: Motion and noise artifact detection," *Ann. Biomed. Eng.*, vol. 42, no. 11, pp. 2238–2250, Nov. 2014.
- [10] Y. Yan, C. C. Poon, and Y. Zhang, "Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner-Ville distribution," *J. NeuroEngineering Rehabil.*, vol. 2, p. 3, Mar. 2005.
- [11] J. Lee, W. Jung, I. Kang, Y. Kim, and G. Lee, "Design of filter to reject motion artifact of pulse oximetry," *Comput. Stand. Interfaces*, vol. 26, no. 3, pp. 241–249, May 2004.
- [12] K. V. P. Narahariseti, M. Bawa, and M. Tahernezahdi, "Comparison of different signal processing methods for reducing artifacts from photoplethysmograph signal," in *2011 IEEE International Conference on Electro/Information Technology (EIT)*, 2011, pp. 1–8.
- [13] M. R. Ram, K. V. Madhav, E. H. Krishna, K. N. Reddy, and K. A. Reddy, "Computation of SpO₂ using non-parametric spectral estimation methods from wavelet based motion artifact reduced PPG signals," in *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, 2011, pp. 776–780.
- [14] R. Krishnan, B. Natarajan, and S. Warren, "Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1867–1876, Aug. 2010.
- [15] J. A. Sukor, S. J. Redmond, and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiol. Meas.*, vol. 32, no. 3, pp. 369–384, Mar. 2011.
- [16] K. Li, 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.*, vol. 6, no. 1, pp. 54–63, Feb. 2012.
- [17] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, Jan. 2008.
- [18] K. Nakajima, T. Tamura, and H. Miike, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique," *Med. Eng. Phys.*, vol. 18, no. 5, pp. 365–372, Jul. 1996.
- [19] 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, no. 10, pp. 1617–1629, Oct. 2012.
- [20] 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, no. 12, p. 1585, Dec. 2010.
- [21] J. Y. A. 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, no. 1–2, pp. 140–145, Mar. 2006.

- [22] R. Krishnan, B. Natarajan, and S. Warren, "Analysis and detection of motion artifact in photoplethysmographic data using higher order statistics," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, 2008, pp. 613–616.
- [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," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 4972–4975, 2011.
- [24] B. Hjorth, "The physical significance of time domain descriptors in EEG analysis," *Electroencephalogr. Clin. Neurophysiol.*, vol. 34, no. 3, pp. 321–325, Mar. 1973.
- [25] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.
- [26] E. Gil, 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*, vol. 3, no. 3, pp. 267–277, Jul. 2008.
- [27] J. W. Chong, 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 detection-reduction approach. Part I: Motion and noise artifact detection," *Ann. Biomed. Eng.*, vol. 42, no. 11, pp. 2238–2250, Nov. 2014.
- [28] C. G. Scully, N. Selvaraj, F. W. Romberg, R. Wardhan, J. Ryan, J. P. Florian, D. G. Silverman, K. H. Shelley, and K. H. Chon, "Using time-frequency analysis of the photoplethysmographic waveform to detect the withdrawal of 900 mL of blood," *Anesth. Analg.*, vol. 115, no. 1, pp. 74–81, Jul. 2012.
- [29] N. Selvaraj, J. Lee, and K. H. Chon, "Time-varying methods for characterizing nonstationary dynamics of physiological systems," *Methods Inf. Med.*, vol. 49, no. 5, pp. 435–442, 2010.
- [30] N. Selvaraj, K. H. Shelley, D. G. Silverman, N. Stachenfeld, N. Galante, J. P. Florian, Y. Mendelson, and K. Chon, "A novel approach using time-frequency analysis of pulse-oximeter data to detect progressive hypovolemia in spontaneously breathing healthy subjects," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, Aug. 2011.
- [31] 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, no. 2, pp. 326–338, Feb. 2006.
- [32] C. . Hsu and C. Chang, "A Practical Guide to Support Vector Classification." Department of Computer Science, National Taiwan University, 2003.
- [33] M. Kearns and D. Ron, "Algorithmic stability and sanity-check bounds for leave-one-out cross-validation," *Neural Comput.*, vol. 11, no. 6, pp. 1427–1453, Aug. 1999.

1. Title page

**Towards the Development of a Mobile Phonopneumogram:
Automatic Breath-Phase Classification using Smartphones**

Bersain A. Reyes¹, Natasa Reljin¹, Youngsun Kong², Yunyoung Nam², Sangho Ha²,
and Ki H. Chon^{1,*}

1 Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269
USA.

2 Department of Computer Science and Engineering, Soonchunhyang University, Asan
336-745 Korea.

* Corresponding author; E-Mail: kchon@engr.uconn.edu; Tel.: +1-860-486-4767.

2. Abstract and key terms

Abstract: Correct labeling of breath phases is useful in the automatic analysis of respiratory sounds, where airflow or volume signals are commonly used as temporal reference. However, such signals are not always available. The development of a smartphone-based respiratory sound analysis system has received increased attention. In this study, we propose an optical approach that takes advantage of a smartphone's camera and provides a chest movement signal useful for classification of the breath phases when simultaneously recording tracheal sounds. Spirometer and smartphone-based signals were acquired from $N=13$ healthy volunteers breathing at different frequencies, airflow and volume levels. We found that the smartphone-acquired chest movement signal was highly correlated with reference volume ($\rho=0.960\pm0.025$, mean $\pm$ SD). A simple linear regression on the chest signal was used to label the breath phases according to the slope between consecutive onsets. 100% accuracy was found for the classification of the analyzed breath phases. We found that the proposed classification scheme can be used to correctly classify breath phases in more challenging breathing patterns, such as those that include non-breath events like swallowing, talking, and coughing, and alternative or irregular breathing. These results show the feasibility of developing a portable and inexpensive phonopneumogram for the analysis of respiratory sounds based on smartphones.

Key Terms: breath-phase classification; respiration; smartphone; smartphone video camera; tracheal sounds; chest movements; phonopneumogram.

3. Introduction

Computerized Respiratory Sound Analysis (CORSAs) has overcome some limitations of the mechanical stethoscope and accelerated the interest in respiratory sound analysis over the last decades³⁹. For example, employment of CORSA systems allows quantification of changes in respiratory sound characteristics, correlation of these sounds to other physiological signals, and generation of data representations useful in the diagnosis and treatment of patients with pulmonary diseases⁷. Even with these advantages, pulmonary auscultation with the stethoscope still guides in diagnosis when other tests are not available²⁷. Ubiquity, low-cost, mobility, ease-of-use, and non-invasiveness are some characteristics that made the stethoscope the most widely used instrument in clinical practice. Such characteristics should remain when aiming for the development of a CORSA system.

The advanced state-of-the-art of smartphones and their near-ubiquity make them an attractive option for developing a CORSA system that provides more useful information than the stethoscope. Employment of smartphones has advantages over other architectures in terms of implementation and integration with other health monitoring technologies given their hardware and software capabilities. Nowadays, smartphone vital sign applications have been found to be accurate and robust in areas such as cardiac and respiratory monitoring^{17,23}.

Automatic classification of breath phases, *i.e.*, automatic labeling of a breath phase as inspiration or expiration, attracts particular interest in applications requiring the timing of breath phases, *e.g.* when studying the breathing modulation of flow in the heart⁴⁴, or during acoustical airflow⁴⁵ and volume estimation³² to correctly assign the polarity of the estimated signals.

In the field of respiratory sounds, discriminating between inspiratory and expiratory phases is also important when analyzing breathing (base) sounds as well as adventitious sounds. The

timing of crackle sounds –short duration (discontinuous) with an explosive character³⁸- must be characterized and it has been found to differ between different pulmonary disorders, reflecting different pathophysiology²⁹. For example, late inspiratory crackles have been associated with restrictive pulmonary diseases while early inspiratory crackles with severe airway obstruction²⁴; early timing of crackles in COPD was found not to overlap with late inspiratory crackles in fibrosing alveolitis³⁰. Expiratory crackles can be found in many respiratory diseases²⁹, *e.g.* low-frequency expiratory crackles occur especially in chronic airway obstruction, but in general they are less frequent than inspiratory crackles⁴³. Similarly, the relationship of continuous adventitious sounds such as wheezes –long duration sounds with a musical character³⁸- to the breath phase is useful for their characterization²⁰. The severity of bronchial obstruction has been found to be less in asthmatic patients with only expiratory wheezes than in patients with both inspiratory and expiratory wheezes³⁶. Inspiratory short duration wheezes (squawks) are commonly heard in pulmonary fibrosing diseases and pneumonia^{8,26}. Regarding base lung sounds, statistically-significant differences were found between healthy and extrinsic allergic alveolitis patients⁵, where the differences were more consistent during the expiratory phase presumably due to the more central source of the expiratory sounds that could carry out more information. Classically, by using phonopneumography –simultaneous presentation of respiratory sound and airflow or volume signals- the timing or volume level of occurrences of adventitious sounds and breath phases can be performed accurately²⁹. However, outside clinical and research settings these airflow or volume signals cannot always be taken for granted.

The idea of developing a portable system for respiratory sound analysis is not new^{10,13}, nor is the idea of using smartphones for such purposes²⁵. Recently, our research group also proposed a smartphone-based system for tracheal sound acquisition purposes³⁴. That study was intended to

show that smartphones allow acquisition of tracheal sounds that resemble the main characteristics reported in the classical literature^{3,14,19,28,37}, such as temporal intensity variation that correlates with airflow, similar frequency content of breath phases at similar airflow peaks, and their use for breath-phase onset detection and respiratory rate estimation. We analyzed the acquired sounds employing a Shannon entropy estimator together with a joint time-frequency technique in order to obtain time-varying respiration rate estimates, which were found to correlate well when compared to reference values from spirometer-acquired signals³⁴. The breath-phase onset estimates based on smartphone-acquired tracheal sounds were found to be around 52 ± 51 ms (mean $\pm$ SD), which are adequate for research involving heart function coupled to respiration⁴⁴. Automatic breath-phase classification was not performed in that previous study.

Use of tracheal sound measurements for estimating ventilation parameters is of particular interest in the CORSA field, *e.g.* phonospirometry provides fairly accurate estimates of airflow⁴⁵ and tidal volume³². Recently, our research group applied a fractal analysis approach for tidal volume estimation from smartphone-acquired tracheal sounds, and it was found that reasonable estimates could be obtained even for measurements five days after calibration using a simple bag at a known volume³³. Besides the promising results in phonospirometry using tracheal sounds, airflow and volume estimators share a necessary step involving the correct classification of the inspiratory and expiratory phases which is usually performed via an additional signal, *e.g.* airflow from a spirometer.

Previous studies using a multichannel CORSA system addressed the classification of breath phases using only respiratory sounds. By employing tracheal sounds for breath-phase onset detection and lung sounds for breath-phase classification, via the inspiratory/expiratory power

difference, even 100% accuracy was achieved²¹. However, the recording of an additional channel was required in order to achieve this. Hence, this former approach is not feasible in a single channel scenario. Its implementation in a smartphone-based CORSA system would require additional hardware to simultaneously acquire two sound channels if intended for tracheal sound analysis. On the other hand, the use of only tracheal sounds for both breath-phase onset detection and breath-phase classification has also been attempted^{1,2,12,14}. By taking advantage of fast changes in tracheal sound intensity, classification has been performed in prior studies using both time and frequency analyses¹². Unfortunately, the accuracy was not reported in the latter case. More recent studies on this classification task have also reported the use of only tracheal sounds, recorded either over the trachea or close to the nostrils or mouth in agreement with current definitions³⁸. By applying a ratio of frequency magnitudes at high and low frequency bands to discriminate between inspiratory and expiratory phases, 97% of 436 phases were correctly classified when compared to respiratory inductance plethysmography². An accuracy of 95.6% was obtained by extracting features from the logarithm of the variance and comparing the current phase to the prior and post phases, with the results being independent of the airflow levels¹⁴. A 90% accuracy for inhalation and exhalation classification was achieved by applying a threshold level to Mel-frequency cepstral coefficients extracted from tracheal sounds¹. As was pointed out by other authors, breath-phase detection is a relatively easy task if lung sounds are used; however, as can be noticed from the reported accuracy results, ranging from 90 to 97%, it is still a topic of ongoing research exploration when employing only tracheal sound recordings. Certainly, there are applications when only recording a single respiratory signal is desirable, and classification of breath phases only from tracheal sounds is advantageous; however, more often other physiological signals are simultaneously recorded in order not only to

enhance the performance of the monitoring system but also to gain a deeper knowledge of the phenomena under analysis.

This study is intended as a step forward towards the development of a mobile CORSA system that takes advantage of smartphone capabilities. Given that smartphones now have a broad collection of sensors, it is natural to question if the employment of additional smartphone-acquired respiratory signals would be helpful when developing a mobile CORSA system. Therefore, as an alternative to the approach of classifying breath phases using only tracheal sounds we propose to acquire an additional respiratory-related signal that can be used as a temporal reference, as it is done in classic phonopneumography, **without the need to plug additional hardware into the smartphone**. In particular, we propose using a smartphone-acquired optical signal that tracks chest movements from which the correct detection of the inspiratory and expiratory phases could be achieved by a simple processing technique directly on the smartphone.

Optical approaches have been used for monitoring cardiac and respiratory parameters^{4,31,41}. Recently, a breathing pattern tracking algorithm was implemented on a personal computer by **detecting shoulder displacements via webcam and image processing techniques³⁵**. In contrast to this study, our research group implemented an application directly on an Android smartphone that recorded chest movements for average respiratory rate estimation²². Similar to the study by **Shao *et al.*³⁵**, we noticed that smartphone-based optical signals resemble the spirometry-based **volume with the uphill and downhill segments corresponding to the inspiratory and expiratory phases**. The proposed smartphone application was previously developed by our research group for non-contact respiratory rate estimation²², and this study is an extension to that work which now intends to perform automatic breath-phase classification for respiratory sound analysis.

Here, as a reference to compare the classification results, spirometer-based airflow and volume signals were simultaneously collected with the chest movement signal recorded remotely from the smartphone's camera. Tracheal sounds were also simultaneously acquired via smartphone as proposed in our previous study³⁴ during noise-free recordings and also while the subjects made non-breath noise (swallow, cough, and talk) and performed both regular (alternate phases) and irregular breathing patterns to analyze the performance of the proposed classification method in such scenarios.

4. Materials and Methods

(a) Subjects

Thirteen ($N=13$) healthy and non-smoker volunteers (twelve males), ages ranging from 19 to 52 years (27.77 ± 9.41 , mean $\pm$ SD), weights 70.77 ± 8.39 kg, and heights 175.31 ± 6.28 cm, were recruited for this study. Students and staff members from the University of Connecticut (UConn), USA, constituted the group of volunteers. Subjects with previous pneumothorax, with chronic respiratory illnesses such as asthma, and anyone who was currently ill (*e.g.* common cold or upper respiratory infection) were excluded from participation. The Institutional Review Board of UConn approved the study protocol which was provided to each volunteer for his/her agreement and signature.

(b) Respiration signals acquisition

(b.1) Equipment and chest movement algorithm

Three types of signals were recorded during the breathing maneuvers of each volunteer: airflow and volume signals via a spirometer, chest movement signals via a smartphone video camera, and tracheal sounds via an acoustical sensor plugged into a smartphone audio input. The spirometer system used for recording the respiratory airflow, and corresponding volume via integration over time, consisted of a respiration flow head connected to a differential pressure transducer (MLT1000L, FE141 Spirometer, ADInstruments, Inc., Dunedin, New Zealand). A 16-bit A/D converter (PowerLab/4SP, ADInstruments, Inc.) was used to sample the analog airflow and volume signals at 1 kHz. Each volunteer received a new disposable filter, reusable mouthpiece, and disposable nose clip compatible with the spirometer system (MLA304, MLA1026, MLA1008, ADInstruments, Inc.). Prior to each volunteer's experiment, the

spirometer system was calibrated using a 3.0 liter calibration syringe (Hans Rudolph, Inc., KS, USA), following instructions in the manufacturer's manual. The digitized volume signal was regarded as a reference for breath-phase classification.

At the same time that the airflow and volume signals were being recorded, each volunteer's chest movement signal was also recorded, using the frontal camera of an HTC One M8 smartphone (HTC Corporation, Taiwan), which consisted of a 5 MP camera with 1080p full HD video recording at 30 frames-per-second (fps) and wide-angle lens. An algorithm was implemented in the smartphone by our research group that recorded chest wall motion at a sampling frequency of 25 Hz during the volunteer's maneuvers²².

It has been shown that during breathing, as in all mechanical systems involving volume displacement, a relationship between volume displacement and linear motion exists, where the rib cage and abdomen compartments of the chest wall are the major contributors¹⁶. Chest wall movements in the anteroposterior direction are greater than those in the vertical or transverse directions, with an increase of around 3 cm in the anteroposterior diameter over the vital capacity range¹⁶. In optical non-contact monitoring of breathing, a video camera captures the changes in the intensity of reflected light caused by these chest wall movements as they modify the path length of the illumination light⁴⁶. The algorithm implemented by our research group averages the intensities of the red, green and blue (RGB) channels of the video within a rectangular region of interest (ROI) at each time instant t as follows

$$I(t) = \left(\frac{1}{3D} \right) \left(\sum_{\{m,n\} \in \text{ROI}} i_R(m,n,t) + \sum_{\{m,n\} \in \text{ROI}} i_G(m,n,t) + \sum_{\{m,n\} \in \text{ROI}} i_B(m,n,t) \right) \quad (1)$$

where D refers to the number of pixels in the ROI, and $i_x(m,n,t)$ refers to the intensity value of the pixel at the m -th row and n -th column of the ROI for the corresponding RGB channel. The ROI was focused on the rib cage area of the subject and consisted of 49×90 pixels selected, i.e.,

$D=4410$ pixels, in a resolution of 320×240 pixels. In order to obtain the chest movement signal $I(t)$, the video data was first converted in the smartphone from YUV420SP format to RGB using the Open Source Computer Vision library⁴⁷. The implemented app saved the recorded chest movement signal $I(t)$ and time vector of the maneuvers in a text file for further analysis in Matlab (R2012a, The Mathworks, Inc., MA, USA).

A Galaxy S4 smartphone (Samsung Electronics Co., Seoul, South Korea) was employed to acquire tracheal sounds via a cabled acoustical sensor composed of a subminiature electret microphone BT-21759-000 (Knowles Electronics, IL, USA) encased in a plastic bell. A double-sided adhesive ring (BIOPAC Systems, CA, USA) was used to affix the acoustical sensor to the volunteers' necks, at the level of the anterior cervical triangle. The Galaxy S4, as well as the HTC One, was running on Android v4.4.2 (KitKat) operating system. The acoustical sensor used in this study was developed by our colleagues at the Metropolitan Autonomous University at Mexico City, and has been successfully used in respiratory sound analysis⁵. The minimum requirements recommended by the European Respiratory Society Task Force Report⁷ are satisfied by the Galaxy S4 high-fidelity audio system, and we found that the characteristics and information that can be extracted from this kind of smartphone-acquired sound signal are in agreement with those using regular CORSA systems³⁴. After smartphone acquisition of the tracheal sounds at 44.1 kHz and 16-bit per sample, the recorded audio files were transferred to a personal computer for further processing in Matlab.

(b.2) Maneuver

Each volunteer was asked to breathe through the spirometer system at airflow levels ranging from around 0.5 to 2 L/s, first increasing their volumetric flow rates with each breath for around

1 minute, and then decreasing volumetric flow rates with each breath for another minute. These airflow levels cover similar ranges as the ones used in other studies when acquiring tracheal sounds at ‘low’, ‘medium’, and ‘high’ airflows^{14,21,32}. Precise minimum and maximum peak airflows varied between volunteers depending on their own manageable levels. For alignment purposes between the different types of recordings, volunteers were asked to perform initial inspiratory and final expiratory apneas of approximately 5 s each and to take a forced respiratory cycle after initial apnea before performing the described maneuver. The airflow signal from the spirometer was displayed on a 40” monitor placed in front of the volunteers to provide them with visual feedback. During the maneuver, volunteers were in standing still posture and wore nose clips to clamp their nostrils. In order to record the chest movement signal, the smartphone was held in a 3-pronged clamp placed in front of the volunteers at approximately 60 cm from their thorax level so that the central portion of their rib cage areas was captured by the ROI. In a real-world application, the distance from the camera to the subject’s thorax would be affected by their body proportions, so it would be necessary to ensure that the ROI’s vertical borders do not exceed the anterior axillary line. We have found that a reliable chest movement signal could be obtained even when the ROI captures a smaller area than that defined by the midclavicular lines. Experiments were performed in a regular dry laboratory, not an anechoic chamber, illuminated with ordinary fluorescent ceiling lights. The laboratory was held quiet during each volunteer’s maneuvers. Volunteers were asked not to wear loose clothes but they were free to wear any pattern, *e.g.*, plain or stripes, and any color of clothing during the maneuvers. Figure 1 shows an example of the setup during a maneuver acquisition.

(c) Data preprocessing

Airflow and volume signals from the spirometer were down-sampled to 25 Hz, and then lowpass filtered at 2 Hz with a 4th-order Butterworth filter applied in a forward and backward scheme to produce zero-phase distortion and minimize start and end transients. Due to fluctuations around the sampling frequency encountered during data acquisition, the chest movement signal was interpolated at 25 Hz via a cubic spline algorithm to obtain a fixed sampling rate. The same lowpass filter at 2 Hz applied to spirometer signals was applied to the chest movement signal to minimize high frequency components not related to the respiratory maneuver. Acquired tracheal sounds were down-sampled to 6300 Hz. To minimize heart sounds and muscle interference, the down-sampled tracheal sounds were filtered using a 4th-order Butterworth bandpass filter between 100 to 3000 Hz and applied in a forward and backward scheme.

Due to differences in starting times and delays between the spirometer system and the smartphones, alignment of smartphone-acquired signals was performed with respect to spirometry. For the chest movement signal, a segment of 20 seconds duration was extracted from each recording at the central portion of the maneuver. The cross-correlation sequence between volume and chest movement segments was computed and the sample lag for which the cross-correlation value resulted in a maximum was used to shift the smartphone-acquired signal accordingly. For the alignment of tracheal sounds, the Shannon entropy (SE) signal was employed as it resembles a rectified version of the airflow signal⁴⁵, with the breath-phase onsets being indicated by its minima. SE was computed in a moving window scheme via the Parzen's density estimation method with a Gaussian kernel⁶ using the parameters detailed in our previous

study³⁴. Then, the tracheal sound was shifted in time so that its initial breath-phase onset after apnea, computed from SE, matched the corresponding onset from the reference volume signal.

Although the manufacturer's instructions were followed, we found a drift in the spirometer-based volume signals. A drift was also found in the smartphone-acquired chest movement signals. Hence, a detrending step based on the Empirical Mode Decomposition (EMD) was applied to both types of signals in order to facilitate their further analysis^{9,11}. EMD employs a sifting process that decomposes the original signal in terms of its intrinsic oscillatory modes (IMFs), based only on the original signal, by analyzing the different time scales presented in it. After the sifting process, the original signal $s(t)$ can be represented as

$$s(t) = \sum_{k=1}^K IMF_k(t) + r_K(t) \quad (2)$$

where K is the total number of IMFs, and $r_K(t)$ is the residual signal. The EMD sifting process is intended to obtain IMFs without riding waveforms and to produce close to zero mean value as defined by their upper and lower envelope signals¹¹. As a result of the sifting process, the first IMFs contain the higher frequency components (lower scales), and hence the trend is contained in the last IMFs. Figure 2 shows an example of raw signals acquired using smartphone and spirometer systems during the breathing maneuver of a volunteer. Observe that even with the baseline drift, the inspiratory/expiratory phases can be noticed as the local increasing/decreasing segments in both the chest movement and reference volume signals. However, signal detrending as done here with EMD, or with a more conventional high-pass digital filter, simplifies the further processing including the automatic onset detection. An example of the preprocessing results is shown in Figure 3.

(d) *Breath phase classification using smartphone camera signals*

As a reference to test the performance of the proposed breath-phase classification, the spirometer's volume signal was used to obtain the actual breath phases during the maneuver. First, the corresponding breath-phase onsets were found via its local maxima and minima. Then, the breath phase between two consecutive onsets was labeled as inspiration or expiration in accordance to the sign of a linear least-squares model¹⁸ fitted on the volume data in that segment (positive: inspiratory phase, negative: expiratory phase).

For the automatic classification of the breath phases using the smartphone-acquired chest movement signal, we propose to take advantage of the linear correlation between the detrended chest movement and the spirometer-based volume signals. As the basis of the proposed method is that the chest movement signal from a smartphone's camera and the spirometer-based volume signal are highly correlated, we quantify this linear correlation during the breathing maneuver by computing the cross-correlation index ρ , defined as:

$$\rho = \frac{\sum_{i=1}^P chest_{smartphone}(i) \cdot volume_{spirometer}(i)}{\sqrt{\sum_{i=1}^P (chest_{smartphone}(i))^2 \cdot \sum_{i=1}^P (volume_{spirometer}(i))^2}} \quad (3)$$

where $chest_{smartphone}$ denotes the smartphone-acquired chest movement signal, $volume_{spirometer}$ the spirometer-acquired volume signal, and P is the total number of samples of the analyzed signals. If both signals were the same, ρ would equal unity. Hence, values close to 1 indicate high correlation between the signals under analysis. Note that if a high linear correlation between smartphone-acquired chest movement and the reference volume signal is found, it would imply that we could easily obtain accurate breath-phase labels from only the chest movement signal. To this end, the chest movement signal was processed in the same way as the volume signal, i.e., the breath-phase onsets were automatically found in the chest

movement signal, then each segment between two consecutive onsets was labeled as inspiration if the sign of the linear least-squares model fitted on the chest movement signal was positive, or as expiration if the corresponding sign was negative, *i.e.*,

$$\text{Breath phase} = \begin{cases} \text{Inspiration,} & \text{if } \text{sign}\{\beta\} > 0 \\ \text{Expiration,} & \text{if } \text{sign}\{\beta\} < 0 \end{cases} \quad (4)$$

where $\text{sign}\{\cdot\}$ refers to the sign function, and β corresponds to the slope of the regression line for the corresponding segment of smartphone data under analysis. For simplicity of notation, let us consider that for every two consecutive breath-phase onsets we have a set of M pairs of smartphone data points denoted by $\{(t_m, y_m)\}_{m=1, \dots, M}$, where $\{y_m\}_{m=1, \dots, M}$ refers to the chest movement data from a smartphone, and $\{t_m\}_{m=1, \dots, M}$ refers to their corresponding time locations at a uniform sampling rate f_s , hence the best linear fit in the least-squares sense has the form $y = \beta t + \alpha$, where the slope β is given by¹⁸

$$\beta = \frac{\sum_{m=1}^M (t_m \cdot y_m) - \frac{1}{M} (\sum_{m=1}^M t_m) \cdot (\sum_{m=1}^M y_m)}{\sum_{m=1}^M (t_m^2) - \frac{1}{M} (\sum_{m=1}^M t_m)^2} \quad (5)$$

Without loss of generality, the relationship between the equidistant time points and the sampling frequency can be used, *i.e.*, $t_m = m \cdot \frac{1}{f_s}$ for $m = 1, \dots, M$ sample indexes, to rewrite the slope of the linear fit as

$$\beta = \frac{\frac{1}{f_s} \sum_{m=1}^M (m \cdot y_m) - \frac{1}{M f_s} (\sum_{m=1}^M m) \cdot (\sum_{m=1}^M y_m)}{\frac{1}{f_s^2} \sum_{m=1}^M m^2 - \frac{1}{M f_s^2} (\sum_{m=1}^M m)^2} \quad (6)$$

Either Equation (5) or (6) could be used for breath-phase classification purposes. However, as our interest is only in the sign of the slope it would be more convenient to reduce computational burden when implemented on the smartphone. Using the closed forms of the finite summations given by

$$\begin{aligned}\sum_{m=1}^M m &= \frac{M(M+1)}{2} \\ \sum_{m=1}^M m^2 &= \frac{M(M+1)(2M+1)}{6}\end{aligned}\tag{7}$$

the Equation of the slope β could be simplified as follows

$$\beta = \left(\frac{6fs}{M(M-1)}\right) \left(\left(\frac{2}{M+1}\right) \sum_{m=1}^M (m \cdot y_m) - \sum_{m=1}^M y_m \right)\tag{8}$$

In turn, by recognizing that in our case the first term in Equation (8) is always positive, the sign of the slope β can be easily computed by

$$\text{sign}\{\beta\} = \text{sign}\left\{ \frac{2 \sum_{m=1}^M (m \cdot y_m)}{M+1} - \sum_{m=1}^M y_m \right\}\tag{9}$$

Finally, the results of the proposed classification scheme using the smartphone-acquired signal can be expressed in terms of the confusion matrix, where the columns are the actual breath-phases as obtained from spirometry, and the rows are the labeled breath-phases from the chest movement signal from smartphone's camera. The accuracy was obtained from the confusion matrix as

$$\text{Accuracy} = \frac{TP + TN}{P + N}\tag{10}$$

where TP refers to the number of actual inspirations correctly labeled as inspirations, TN to the number of actual expirations correctly labeled as expirations, and P and N to the total number of actual inspirations and expirations, respectively.

5. Results

Table 1 contains statistics about breath-phase duration, peak airflow, and tidal volume for the breathing maneuvers performed by $N=13$ volunteers, as measured from spirometer-based airflow and volume signals. The analyzed database was composed of $n_1=419$ inspirations and $n_2=430$ expirations.

The smartphone-acquired chest movement signal follows the temporal variations of the spirometer-based volume signal during the breathing maneuvers, as shown from the raw data in Figure 2 and more clearly in Figure 3 after alignment and detrending. We found a high linear relationship between both detrended signals for all volunteers as measured by the cross-correlation index, $\rho=0.960\pm0.025$. Figure 4 shows an example of the proposed method for automatic breath-phase classification using the smartphone-acquired chest movement signal. Table 2 presents the classification results of the breath phases, as a confusion matrix, for all breathing phases performed by volunteers, where the actual breath phases were obtained from spirometer-acquired volume signals. 100% classification accuracy was achieved as can be seen from the confusion matrix shown in Table 2.

In addition to the previous breathing maneuvers, a couple of volunteers were asked to perform additional breathing patterns according to different scenarios plausible to occur during respiratory recordings, as has been pointed out¹⁴. Additional recordings included the following scenarios: non-breath noise immersed in regular or irregular breathing, and successive inhalations or exhalations. The scenario with alternating breathing phases with different durations (inspiration-expiration-inspiration-expiration) was not explicitly performed at this time because it was already achieved during the main breathing maneuvers performed by all volunteers. At this stage of the study, the chest movement algorithm was already implemented

on the Samsung S4 smartphone so that only this device was employed for both tracheal sounds and chest movements recording. The Samsung S4 frontal camera -2 MP, 1080p video recording @ 30 fps- was employed for chest movement recording. As before, the native resolution of the Samsung S4 device was not used due to computational burden; its resolution was reduced to 320×240 pixels and the ROI was set to 49×90 pixels to match those parameters used in the HTC One smartphone. Examples of recorded signals from two volunteers performing different breathing scenarios with non-breathing noises are shown in Figures 5 and 6. Examples of signals acquired while the volunteers breathed in successive phases are presented in Figure 7. In Figures 5-7, airflow and volume signals are displayed for temporal reference; gray and black bars displayed on top indicate the inspiratory and expiratory phases, respectively. Fitted lines are superimposed on chest movement signals from the smartphone to show the phase labeling outside the noise event as determined by the corresponding slopes. In Figures 5 and 6, the noise events are indicated by a red bar. These events were labeled by examining the sound replay and waveform display of the tracheal sounds simultaneously with the chest movement signal from the smartphone, similar to the common practice in respiratory sound analysis, *e.g.* when labeling adventitious sound events using phonopneumography. Observe that in these cases, the classification of the breath phases is concerned with the phases surrounding the noise events. In Figure 7, the occurrence of successive inspirations and expirations are also indicated by a red bar, where classification of these breath phases is of concern. By employing the slope of the smartphone signal, these successive phases will be correctly classified with the same phase label given the monotonically increasing (or decreasing) chest movement waveform in such segments.

6. Discussion

In this paper we propose the automatic classification of inspiratory and expiratory phases from a smartphone-acquired optical recording as an extension to the acquisition of tracheal sounds via smartphones. The app we developed allowed real time recording of chest movements during breathing maneuvers directly on the smartphone. For this study, the app was implemented on two Android smartphones, the HTC One M8 and the Samsung Galaxy S4. During the initial stage of the study, recordings of chest movements and tracheal sounds were obtained on two separate smartphones, *i.e.*, the HTC One recorded chest movements and the Galaxy S4 recorded tracheal sounds, as each corresponding smartphone was proposed for that particular use in our previous studies^{22,34}. In the second stage of this study, both types of recordings were performed on the same smartphone, *i.e.*, the Galaxy S4 simultaneously recorded chest movements and tracheal sounds.

Previously we studied the employment of smartphones for developing a CORSA system³⁴. Results found in that study motivated us to keep working toward the development of a low-cost, easy-to-upgrade, and reliable portable CORSA system. In a subsequent study, our research group aimed for tidal volume estimation using smartphone-acquired tracheal sounds together with novel signal processing techniques and a simple calibration method that does not involve expensive or specialized devices such as spirometers³³. Although the results are promising, the proposed methods require the correct identification of the inspiratory and expiratory phases.

Phonopneumography has been useful in the field of respiratory sound analysis. When available, it is used as temporal reference for detection and classification of breath phases as well as diverse time events occurring during the breathing maneuver. **Accordingly, the correct classification of breath phases proves to be relevant when performing automatic analysis of**

respiratory sounds containing adventitious sounds^{20,29}, as well as for applications involving airflow or volume estimation from tracheal sounds^{32,44,45}. Given the promising estimation of ventilation parameters, the use of only tracheal sounds has been proposed to address the automatic classification of breath phases^{1,2,12,14}. Although this approach has advantages, *e.g.*, greater user acceptance of the acoustical sensors in comparison to nasal cannulas or facemasks used to measure airflow, its accuracy results for breath-phase classification have not matched those found when using an additional lung sound channel.

Given the importance of the correct breath-phase classification in the CORSA field, and as a more-accurate alternative to using only tracheal sounds, we studied the employment of an additional respiration-related signal that could easily upgrade a mobile smartphone-based system. In this paper, instead of attempting the classification of breath phases from tracheal sounds, we employed an optical approach to perform this task. Previously, our research group implemented an algorithm that allows the estimation of average respiratory rate from a smartphone-acquired chest movement signal²², and we noticed that this signal resembles the spirometer-acquired volume signal. To investigate our previous visual observations, in this study we compared the spirometer-based volume and the smartphone-based chest movement signal using the cross-correlation index. The chest movements and tracheal sounds were recorded on separate smartphones at the initial stage of the study because the optical algorithm had been only implemented on a different smartphone from the one used to record tracheal sounds in our previous studies. We found that both types of signals were highly correlated ($\rho=0.960\pm0.025$, mean $\pm$ SD), corroborating our initial observations. These results indicate that our smartphone-based monitor is able to capture the intensity changes in the reflected light caused by the chest motion, linearly related to volume, while breathing. According to Konno and Mead¹⁶, this

motion-volume linear relationship is attributable to the relative smaller diameter changes while breathing in comparison to the absolute diameter of the chest wall, and to the larger contribution of the anteroposterior diameter changes compared to the vertical or transversal. This linearity appears to hold in the recorded optical chest movement signal from a smartphone's camera.

Hence, the volume signal was employed to label the phases of the respiratory maneuvers, while the chest movement signal was processed using a simple linear regression to label the uphill segments as inspirations and downhill segments as expirations based on the slope of the computed model. We found 100% accuracy for the task of breath-phase classification, *i.e.*, all inspiratory phases ($n_1=419$) were detected as inspirations and all expiratory phases ($n_2=430$) were detected as expirations, for the maneuvers performed by the volunteers in standing still posture, while breathing at different cycle durations ranging from 700 milliseconds to 3 seconds, and different airflow levels with peaks ranging from 0.5 to 2.0 L/s.

The second stage of the study was intended to analyze the performance of the chest movement signal for the automatic classification of breath phases during different scenarios of breathing patterns that included non-alternate inspiratory and expiratory phases, as well as non-breathing related noises like swallowing, talking and coughing. At this point, the optical algorithm was already implemented on the same smartphone tested for tracheal sound acquisition, so that in this stage only a single smartphone was employed. As stated by other authors, these different breathing patterns are the most challenging in respiratory phase detection¹⁴. We found that the proposed classification scheme can be used to correctly classify the breath phases in such scenarios. For the non-breath events immersed in typical alternate breathing (*e.g.* inspiration-noise-expiration) or in irregular breathing (*e.g.* expiration-noise-expiration), the algorithm was able to classify the breath phases surrounding these noise events as indicated by the

corresponding slopes of the chest movement signal from the smartphone. During the scenarios involving consecutive inspirations or consecutive expirations, the tracheal sounds involved were correctly classified as the same phase given the fitted slope for the chest movement signal in that time interval.

Besides the above-mentioned results, we recognize limitations of this study. First, subjects were instructed to stand still while performing the breathing maneuvers, and hence, the performance deterioration due to body motion artifacts, not related to the breathing maneuver, was not explored. Incorporation of body tracking and artifact removal algorithms similar to those proposed in the literature to reduce such motion effects –for example in^{35,40}– is a topic of further exploration towards the development of our mobile system. Second, we only explored recordings with the subjects in standing posture. Recordings in supine posture were not performed. We foresee that the proposed scheme would bring similar classification results to the ones reported here when the visual field of the smartphone’s camera is focused on the area with the most dominant contribution to volume while breathing, *e.g.* the abdominal compartment in supine posture¹⁶. Third, recordings were performed in a regular indoor laboratory, and hence further experiments are required to analyze the usability of the proposed portable system in different outdoor environments to fully take advantage of its mobility.

This study represents a step forward in the development of a mobile system for the analysis of respiratory sounds that takes advantage of additional sensors already existing in smartphones. The obtained results show that simultaneous recordings of tracheal sounds and chest movements are useful for both automatic classification of the breath phases and correct timing of events such as the ones shown in this paper. An interesting alternative to our proposed approach and a topic for future exploration involves the use of accelerometers for respiratory sound recording^{15,42} with

the potential benefit that information regarding the breath phase could be extracted, especially for lung sound recordings, in addition to the respiratory sound itself. Currently, motivated by the high linear correlation obtained between the chest movement signal from the smartphone's camera and the reference volume from spirometry, we are working on a study involving the feasibility of estimating tidal volume via the smartphone-acquired chest movement signal so that estimation of this parameter could be easily performed outside research and clinical settings. Finally, we consider that the smartphone approach proposed in this study, as well as similar ones for respiratory monitoring, has the potential to be readily accepted by users due to its simplicity and comfort as well as potential to reach populations and geographic areas where it is difficult to study respiratory sounds with current computerized methods.

7. 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 an acoustical sensor used in this study.

8. References

1. Abushakra, A., and M. Faezipour. Acoustic signal classification of breathing movements to virtually aid breath regulation. *IEEE J. Biomed. Health Inform.* 17:493–500, 2013.
2. Alshaer, H., G. R. Fernie, and T. D. Bradley. Monitoring of breathing phases using a bioacoustic method in healthy awake subjects. *J. Clin. Monit. Comput.* 25:285–294, 2011.
3. Beck, R., G. Rosenhouse, M. Mahagnah, R. M. Chow, D. W. Cugell, and N. Gavriely. Measurements and Theory of Normal Tracheal Breath Sounds. *Ann. Biomed. Eng.* 33:1344–1351, 2005.
4. Cala, S. J., C. M. Kenyon, G. Ferrigno, P. Carnevali, A. Aliverti, A. Pedotti, P. T. Macklem, and D. F. Rochester. Chest wall and lung volume estimation by optical reflectance motion analysis. *J. Appl. Physiol.* 81:2680–2689, 1996.
5. Charleston-Villalobos, S., L. Albuene-Sanchez, R. Gonzalez-Camarena, M. Mejia-Avila, G. Carrillo-Rodriguez, and T. Aljama-Corrales. Linear and Nonlinear Analysis of Base Lung Sound in Extrinsic Allergic Alveolitis Patients in Comparison to Healthy Subjects. *Methods Inf. Med.* 52:266–276, 2013.
6. Duda, R. O., P. E. Hart, and D. G. Stork. Pattern Classification. New York, NY, USA: Wiley-Interscience, 2000, 679 pp.
7. Earis, J. E., and B. M. G. Cheetham. Current methods used for computerized respiratory sound analysis. *Eur. Respir. Rev.* 10:586–590, 2000.
8. Earis, J. E., K. Marsh, M. G. Pearson, and C. M. Ogilvie. The inspiratory “squawk” in extrinsic allergic alveolitis and other pulmonary fibroses. *Thorax* 37:923–926, 1982.
9. Flandrin, P., P. Goncalves, and G. Rilling. Detrending and denoising with empirical mode decomposition. , 2004.
10. Guangbin, L., C. Shaoqin, Z. Jingming, C. Jinzhi, and W. Shengju. The development of a portable breath sounds analysis system. , 1992.
11. Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* 454:903–995, 1998.
12. Hult, P., B. Wranne, and P. Ask. A bioacoustic method for timing of the different phases of the breathing cycle and monitoring of breathing frequency. *Med. Eng. Phys.* 22:425–433, 2000.
13. Hung, K., B. L. Luk, W. H. Choy, B. Tai, and S. K. Tso. Multifunction stethoscope for telemedicine. , 2004.
14. Huq, S., and Z. Moussavi. Acoustic breath-phase detection using tracheal breath sounds. *Med. Biol. Eng. Comput.* 50:297–308, 2012.
15. Hu, Y., E. G. Kim, G. Cao, S. Liu, and Y. Xu. Physiological acoustic sensing based on accelerometers: a survey for mobile healthcare. *Ann. Biomed. Eng.* 42:2264–2277, 2014.
16. Konno, K., and J. Mead. Measurement of the separate volume changes of rib cage and abdomen during breathing. *J. Appl. Physiol.* 22:407–422, 1967.
17. Lee, J., B. A. Reyes, D. D. McManus, O. Mathias, and K. H. Chon. Atrial Fibrillation Detection Using an iPhone 4S. *IEEE Trans. Biomed. Eng.* 60:203–206, 2013.
18. Lindfield, G., and J. Penny. Numerical Methods: Using MATLAB. Academic Press, 2012, 553 pp.

19. McKusick, V. A., J. T. Jenkins, and G. N. Webb. The acoustic basis of the chest examination; studies by means of sound spectrography. *Am. Rev. Tuberc.* 72:12–34, 1955.
20. Meslier, N., G. Charbonneau, and J. L. Racineux. Wheezes. *Eur. Respir. J.* 8:1942–1948, 1995.
21. Moussavi, Z. K., M. T. Leopando, H. Pasterkamp, and D. G. Rempel. Computerised acoustical respiratory phase detection without airflow measurement. *Med. Biol. Eng. Comput.* 38:198–203, 2000.
22. Nam, Y., Y. Kong, B. Reyes, N. Reljin, and K. H. Chon. Monitoring of Heart and Respiratory Rates using Dual Cameras on a Smartphone. In revision:, 2015.
23. Nam, Y., J. Lee, and K. H. Chon. Respiratory Rate Estimation from the Built-in Cameras of Smartphones and Tablets. *Ann. Biomed. Eng.* 42:885–898, 2013.
24. Nath, A. R., and L. H. Capel. Inspiratory crackles—early and late. *Thorax* 29:223–227, 1974.
25. Oletic, D., B. Arsenali, and V. Bilas. Towards Continuous Wheeze Detection Body Sensor Node as a Core of Asthma Monitoring System. In: *Wireless Mobile Communication and Healthcare*, edited by K. S. Nikita, J. C. Lin, D. I. Fotiadis, and M.-T. A. Waldmeyer. Springer Berlin Heidelberg, 2012, pp. 165–172.
26. Paciej, R., A. Vyshedskiy, D. Bana, and R. Murphy. Squawks in pneumonia. *Thorax* 59:177–178, 2004.
27. Pasterkamp, H., C. Carson, D. Daien, and Y. Oh. Digital respirosography. New images of lung sounds. *Chest* 96:1405–1412, 1989.
28. Pasterkamp, H., S. S. Kraman, and G. R. Wodicka. Respiratory Sounds: Advances Beyond the Stethoscope. *Am. J. Respir. Crit. Care Med.* 156:974–987, 1997.
29. Piirila, P., and A. R. Sovijarvi. Crackles: recording, analysis and clinical significance. *Eur. Respir. J.* 8:2139–2148, 1995.
30. Piirilä, P., A. R. Sovijärvi, T. Kaisla, H. M. Rajala, and T. Katila. Crackles in patients with fibrosing alveolitis, bronchiectasis, COPD, and heart failure. *Chest* 99:1076–1083, 1991.
31. Poh, M.-Z., D. J. McDuff, and R. W. Picard. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Trans. Biomed. Eng.* 58:7–11, 2011.
32. Que, C.-L., C. Kolmaga, L.-G. Durand, S. M. Kelly, and P. T. Macklem. Phonspirometry for noninvasive measurement of ventilation: methodology and preliminary results. *J. Appl. Physiol. Bethesda Md* 1985 93:1515–1526, 2002.
33. Reljin, N., B. A. Reyes, and K. H. Chon. Tidal Volume Estimation using Blanket Fractal Dimension of the Tracheal Sounds Acquired by Smartphone. *Sensors* In revision for publication:, 2015.
34. Reyes, B. A., N. Reljin, and K. H. Chon. Tracheal Sounds Acquisition Using Smartphones. *Sensors* 14:13830–13850, 2014.
35. Shao, D., Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao. Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time. *IEEE Trans. Biomed. Eng.* 61:2760–2767, 2014.
36. Shim, C. S., and M. H. Williams. Relationship of wheezing to the severity of obstruction in asthma. *Arch. Intern. Med.* 143:890–892, 1983.
37. Sierra, G., V. Telfort, B. Popov, M. Pelletier, P. Despault, R. Agarwal, and V. Lanzo. Comparison of respiratory rate estimation based on tracheal sounds versus a capnograph. , 2005.

38. Sovijarvi, A. R. A., F. Dalmasso, J. Vanderschoot, L. P. Malmberg, G. Righini, and S. A. T. Stoneman. Definition of terms for applications of respiratory sounds. *Eur. Respir. Rev.* 10:597–610, 2000.
39. Sovijarvi, A. R. A., J. Vanderschoot, and J. E. Earis. Standardization of computerized respiratory sound analysis. *Eur. Respir. Rev.* 10:585–585, 2000.
40. Sun, Y., S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *J. Biomed. Opt.* 16:077010–077010, 2011.
41. Tarassenko, L., M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiol. Meas.* 35:807, 2014.
42. Vannuccini, L., J. E. Earis, P. Helisto, B. M. G. Cheetham, M. Rossi, A. R. A. Sovijarvi, and J. Vanderschoot. Capturing and preprocessing of respiratory sounds. *Eur. Respir. Rev.* 10:616–620, 2000.
43. Vyshedskiy, A., R. M. Alhashem, R. Paciej, M. Ebril, I. Rudman, J. J. Fredberg, and R. Murphy. Mechanism of inspiratory and expiratory crackles. *Chest* 135:156–164, 2009.
44. Xiong, C., B. Hök, T. Strömberg, D. Loyd, B. Wranne, and P. Ask. A bioacoustic method for timing of respiration at cardiac investigations. *Clin. Physiol.* 15:151–157, 1995.
45. Yadollahi, A., and Z. M. K. Moussavi. Acoustical Respiratory Flow. *IEEE Eng. Med. Biol. Mag.* 26:56–61, 2007.
46. Zhao, F., M. Li, Y. Qian, and J. Z. Tsien. Remote Measurements of Heart and Respiration Rates for Telemedicine. *PLoS ONE* 8:e71384, 2013.
47. Open CV. Open Source Computer Vision Library at <http://sourceforge.net/projects/opencvlibrary/>

9. Tables, figures and legends

Figures:

Figure 1, Bersain Reyes, ABME:

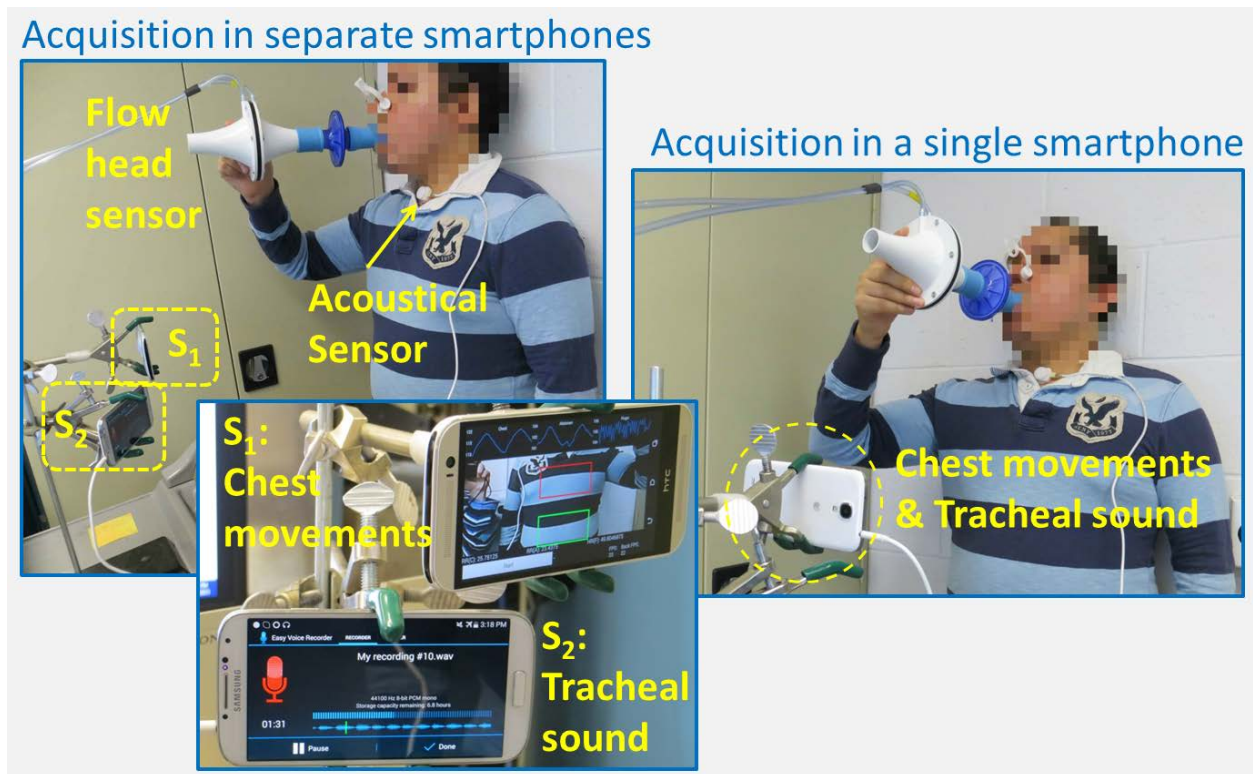


Figure 2, Bersain Reyes, ABME:

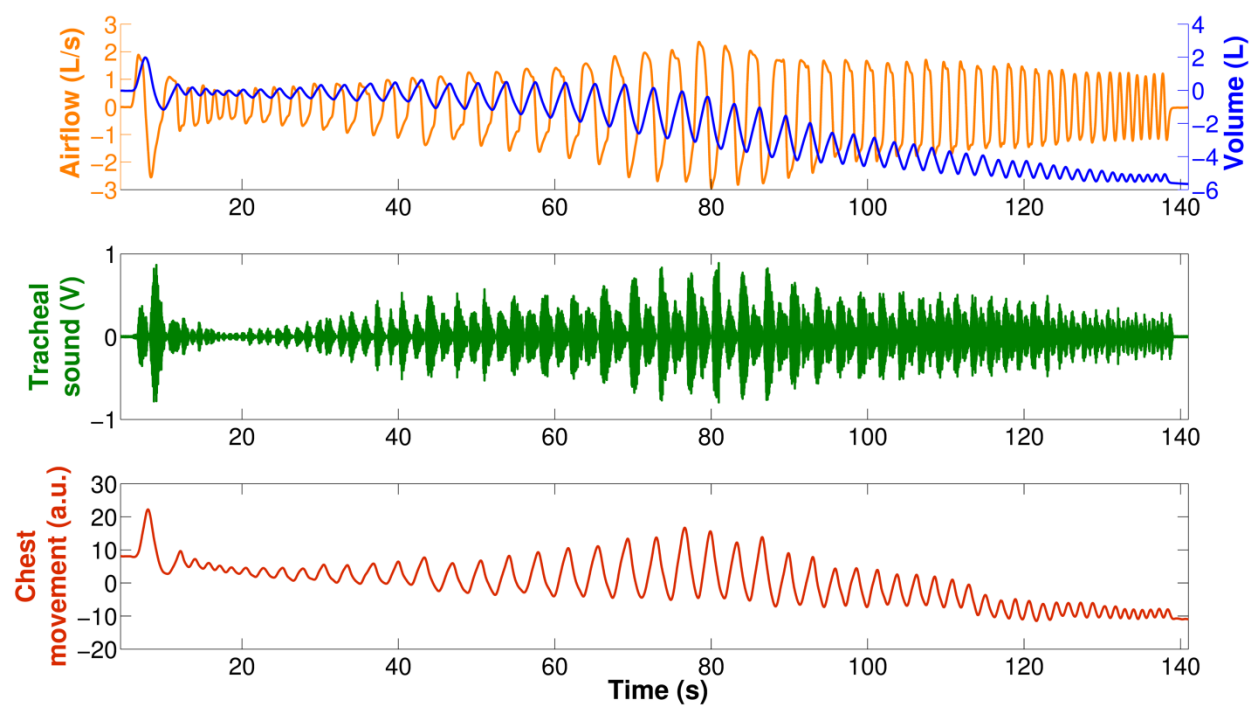


Figure 3, Bersain Reyes, ABME:

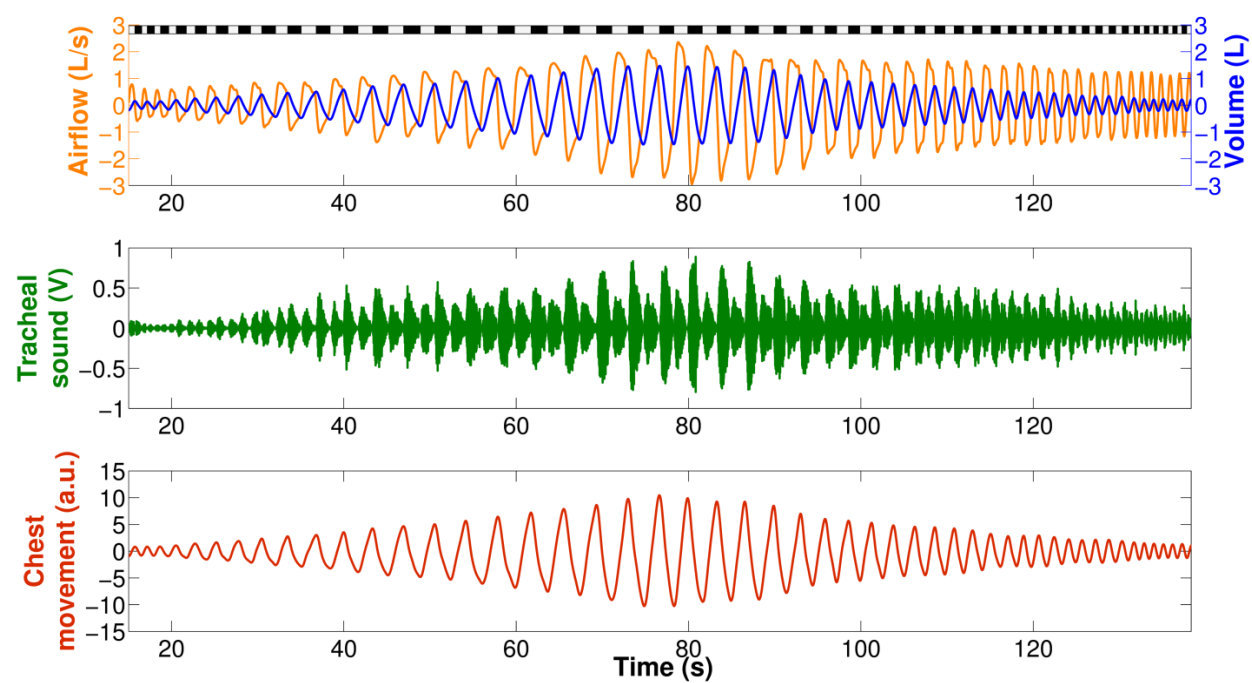


Figure 4, Bersain Reyes, ABME:

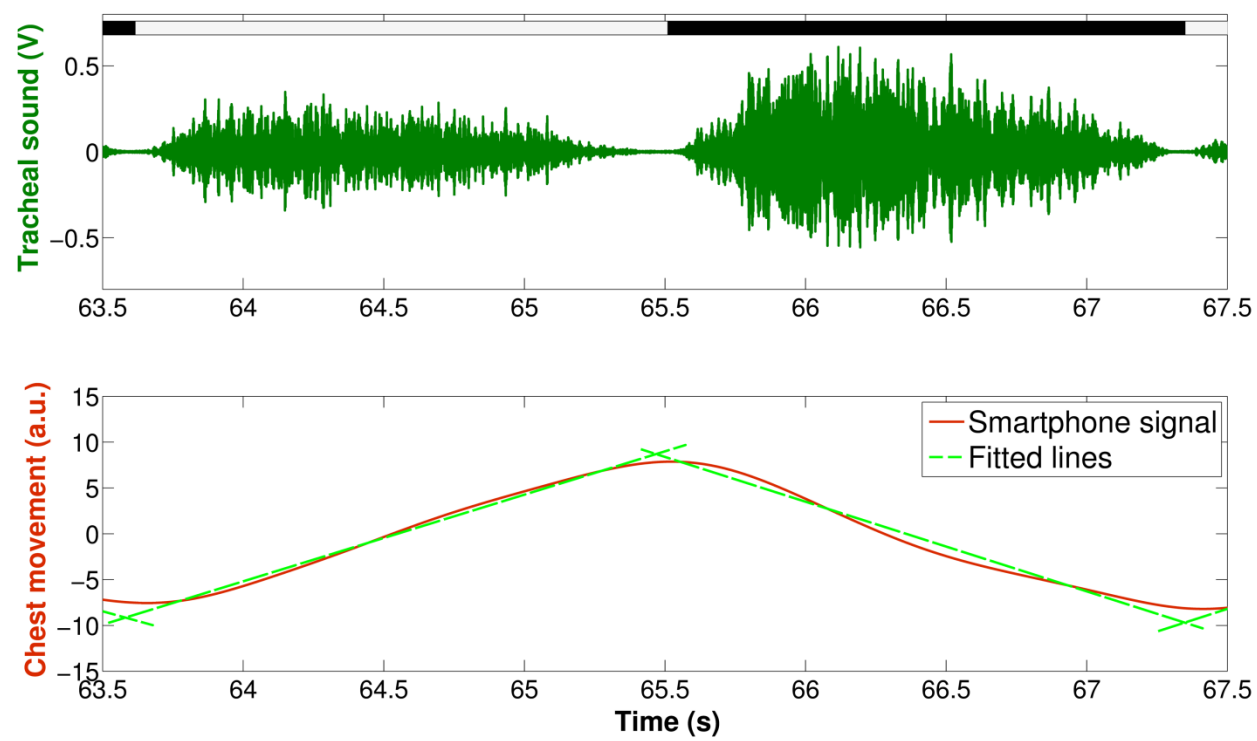


Figure 5, Bersain Reyes, ABME:

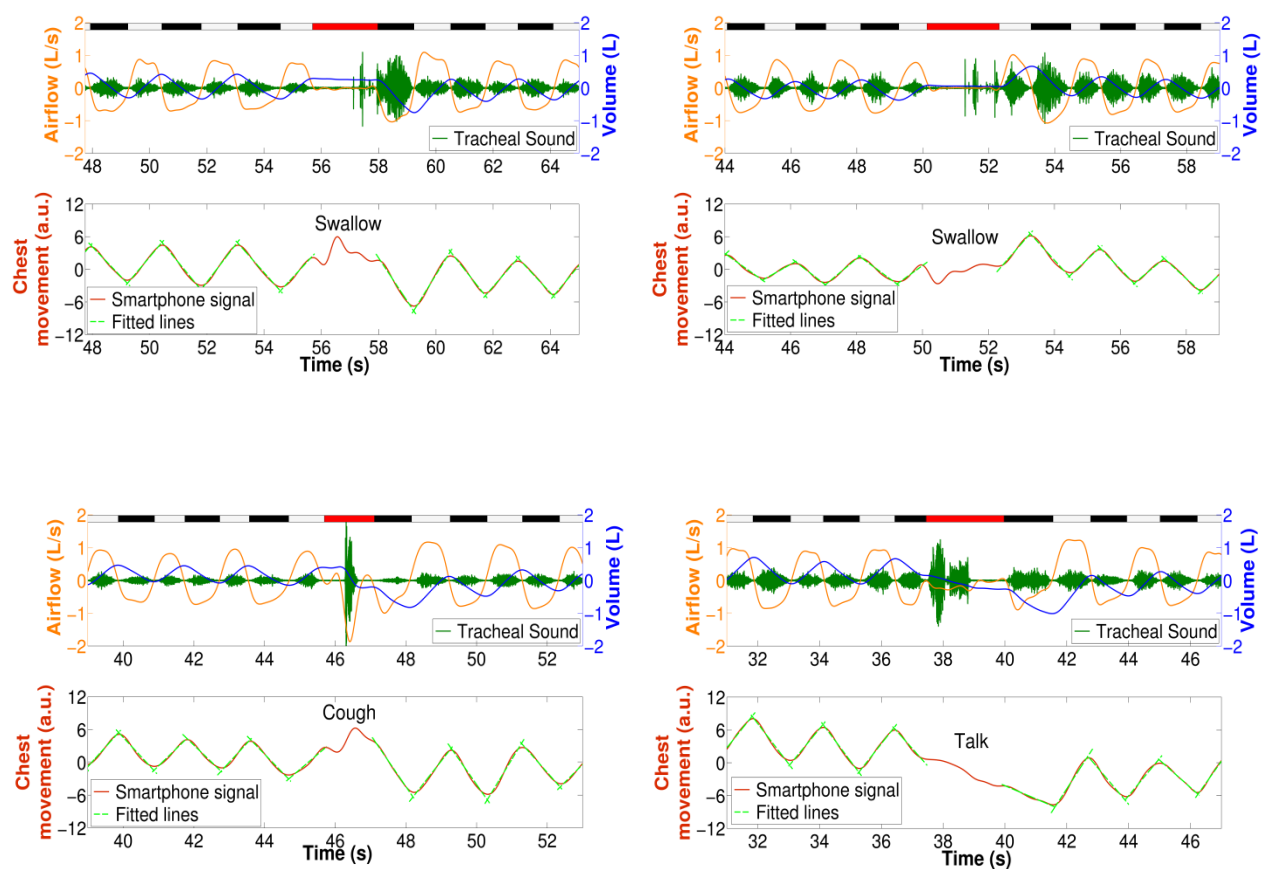


Figure 6, Bersain Reyes, ABME:

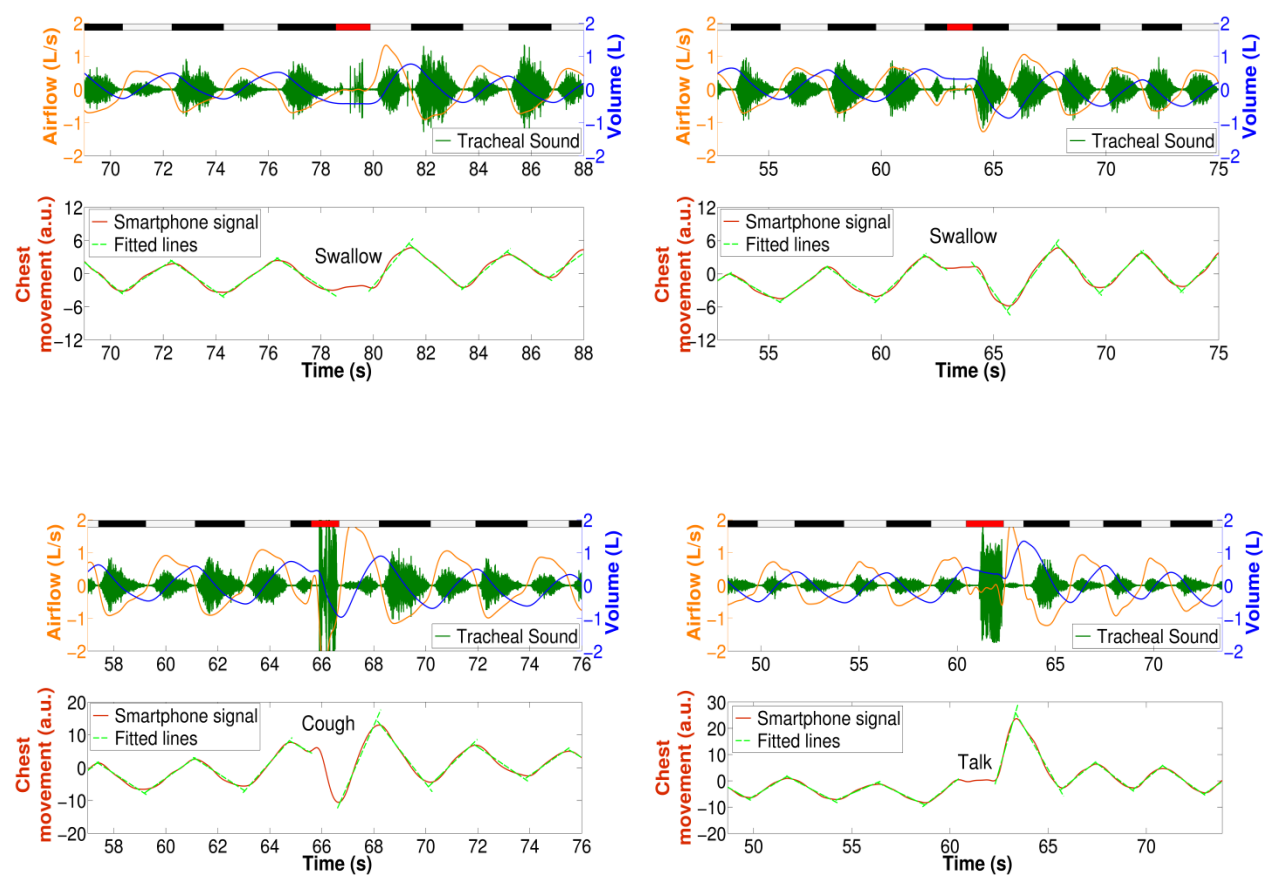


Figure 7, Bersain Reyes, ABME:

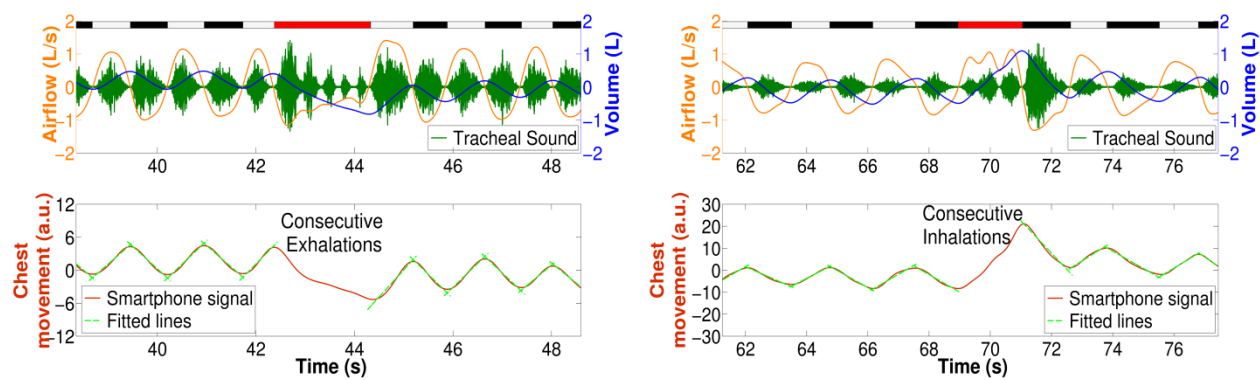


Figure captions:

Figure 1. Recording of breathing signals during the maneuver. A smartphone was placed in front of the volunteer at his/her thorax level in order to record the chest movements directly on this device. Tracheal sounds were acquired with an acoustical sensor plugged into the smartphone. Two separate devices were employed to acquire tracheal sounds and chest movement signals in the first stage of the study. Acquisition of both signals was performed with a single smartphone in the second stage. Airflow and volume signals were also acquired via a spirometer system and regarded as temporal reference. Actual breath-phases of the maneuver were obtained from volume signal.

Figure 2. Example of acquired signals during the respiration maneuver of a volunteer. **Top:** spirometer-acquired airflow (orange) and volume (blue) signals. **Middle:** smartphone-acquired tracheal sounds. **Bottom:** smartphone-acquired chest movement signal. Observe that despite of the baseline drift and different starting times, the breath-phase onsets are noticeable in both reference volume from spirometer and chest movement signal from smartphone's camera.

Figure 3. Example of preprocessed signals during the breathing maneuver of a volunteer. **Top:** spirometer-acquired airflow signal (orange) and volume signal (blue) after detrending. **Middle:** smartphone-acquired tracheal sounds. **Bottom:** smartphone-acquired chest movement signal with the baseline drift removed after detrend. Gray and black bars displayed on top of spirometer signals indicate the inspiratory and expiratory phases, respectively. Both types of

smartphone-acquired signals were aligned in time with respect to reference volume from spirometer.

Figure 4. Example of automatic breath-phase classification using the smartphone-acquired chest movement signal. **Top:** smartphone-acquired tracheal sound signal. Gray and black bars displayed on top indicate the inspiratory and expiratory phases, respectively, as measured from reference volume signal from spirometry. **Bottom:** smartphone-acquired chest movement signal. Superimposed dashed green lines indicate the fitted lines computed via least-squares method. Positive and negative slopes of fitted lines were used to label the segment as inspiration and expiration, respectively.

Figure 5. Example of smartphone-acquired signals during different scenarios of breathing patterns. For each of the four panels, the upper graph displays the airflow (orange), volume (blue), and tracheal sound (dark green) signals, while the bottom graph displays the chest movement signal (red) and the fitted lines computed via least-squares (dashed green lines). Gray/black bars displayed on top indicate the actual inspiratory/expiratory phases measured from spirometry, while the red bar indicates the location of the non-breath noise event. **Top left panel:** non-breath noise event (swallow) immersed in regular breathing patterns. **Top right panel:** non-breath noise event (swallow) immersed in irregular breathing. **Bottom left panel:** non-breath event noise (cough) immersed in regular breathing. **Bottom right panel:** non-breath noise event (talk) immersed in irregular breathing.

Figure 6. Example of smartphone-acquired signals during different scenarios of breathing patterns of a second volunteer. For each of the four panels, the upper graph displays the airflow (orange), volume (blue), and tracheal sound (dark green) signals, while the bottom graph displays the chest movement signal (red) and the fitted lines computed via least-squares (dashed green lines). Gray/black bars displayed on top indicate the actual inspiratory/expiratory phases measured from spirometry, while the red bar indicates the location of the non-breath noise event.

Top left panel: non-breath noise event (swallow) immersed in regular breathing patterns. **Top right panel:** non-breath noise event (swallow) immersed in irregular breathing. **Bottom left panel:** non-breath noise event (cough) immersed in regular breathing. **Bottom right panel:** non-breath noise event (talk) immersed in irregular breathing.

Figure 7. Example of acquired respiratory signals while a couple of volunteers were taking successive breaths. For each of the two panels, the upper graph displays the airflow (orange), volume (blue), and tracheal sound (dark green) signals, while the bottom graph displays the chest movement signal (red) and the fitted lines computed via least-squares (dashed green lines). Gray/black bars displayed on top indicate the actual inspiratory/expiratory phases measured from spirometry, while the red bar indicates the location of the successive breaths event. **Left panel:** consecutive exhalations. **Right panel:** consecutive inhalations.

Tables:

Table 1. Distribution of breath phases' duration, tidal volume, and peak airflow obtained from spirometer during breathing maneuvers ($N=13$ subjects. Number of expirations=430. Number of inspirations=419).

Parameter	Minimum	Maximum	Mean	Median
<i>Phase duration</i> [s]	0.739 ± 0.317	3.211 ± 1.160	1.749 ± 0.586	1.720 ± 0.670
Inspiration				
<i>Peak airflow</i> [L/s]	0.478 ± 0.176	2.232 ± 1.127	1.107 ± 0.286	1.022 ± 0.263
<i>Tidal volume</i> [L]	0.268 ± 0.131	2.986 ± 0.651	1.292 ± 0.222	1.090 ± 0.215
Expiration				
<i>Peak airflow</i> [L/s]	-0.426 ± 0.203	-2.144 ± 0.875	-1.064 ± 0.361	-0.976 ± 0.346
<i>Tidal volume</i> [L]	-2.972 ± 0.683	-0.236 ± 0.114	-1.261 ± 0.213	-1.062 ± 0.225
Values are presented as mean ± standard deviation				

Table 2. Breath-phase classification results using smartphone-acquired chest movement signal ($N=13$ subjects. Number of actual expirations=430. Number of actual inspirations=419).

		Actual Breath Phase (Spirometer)	
		Expiration	Inspiration
Classified Breath phase (Smartphone)	Expiration	430	0
	Inspiration	0	419

Tidal Volume and Instantaneous Respiration Rate Estimation using a Volumetric Surrogate Signal Acquired via a Smartphone Camera

Bersain A. Reyes, *Student Member, IEEE*, Natasa Reljin, Youngsun Kong, *Student Member, IEEE*, Yunyoung Nam, *Member, IEEE*, and Ki H. Chon*, *Senior Member, IEEE*

Abstract—Two parameters that a breathing status monitor should provide include tidal volume (V_T) and respiration rate (RR). Recently we implemented an optical monitoring approach that tracks chest wall movements directly on a smartphone. In this paper, we explore the use of such noncontact optical monitoring to obtain a volumetric surrogate signal, via analysis of intensity changes in the video channels caused by the chest wall movements during breathing, in order to provide not just average RR, but also information about V_T and to track RR at each time-instant (IRR). The algorithm, implemented on an Android smartphone, was used to analyze the video information from the smartphone's camera and provide in real time the chest movement signal from $N=15$ healthy volunteers breathing at V_T ranging from 300 mL to 3 L. Simultaneous recording of volume signals from a spirometer was regarded as reference. A highly linear relationship between peak-to-peak amplitude of the smartphone-acquired chest movement signal and spirometer V_T was found ($r^2=0.951 \pm 0.042$, mean $\pm$ SD). After calibration on a subject-by-subject basis, no statistically-significant bias was found in terms of V_T estimation; the 95% limits of agreement were -0.348 to 0.376 L, and the RMSE was 0.182 ± 0.107 L. In terms of IRR estimation, a highly linear relation between smartphone estimates and the spirometer reference was found ($r^2=0.999 \pm 0.002$). The bias, 95% limits of agreement, and RMSE were -0.024 bpm, -0.850 to 0.802 bpm, and 0.414 ± 0.178 bpm, respectively. These promising results show the feasibility of developing an inexpensive and portable breathing monitor which could provide information about IRR as well as V_T , when calibrated on an individual basis, using smartphones. Further studies are required to enable practical implementation of the proposed approach.

Index Terms—tidal volume, respiration rate, volume surrogate, smartphone camera, optical monitoring, time-frequency analysis.

I. INTRODUCTION

Monitoring of respiration status has been recognized as critical to identifying and predicting serious adverse events [1], [2]. Two basic parameters that a breathing monitor should be able to provide are tidal volume (V_T) and respiration rate (RR) [3]. V_T provides information about the respiration

depth and is defined as the volume of air moved with each breath; on the other hand, RR corresponds to the number of breaths per unit of time and it is commonly expressed in breaths-per-minute. In turn, the product of these two quantities defines the volume of gas moved by the respiratory system per minute, called minute ventilation ($\dot{V}_E$). Normal average values for a human man are around 0.5 L and 12 breaths-per-minute (bpm) for V_T and RR, respectively. These values are not fixed and the mechanism of respiratory control is crucial in determining $\dot{V}_E$ by adjusting the combination of V_T and RR according to a body's requirements in response to different scenarios [4].

Current clinical continuous RR monitoring methods include qualified human observation, transthoracic impedance, inductance plethysmography, capnography monitoring, and tracheal sound monitoring [5]–[8]. Each method has its own disadvantages, e.g. it is time consuming and subjective to do human observation, patients have a low tolerance for using the nasal cannula in capnography [3]. However flawed, at least clinical devices exist for monitoring. Outside clinical or research settings, there is still a lack of monitoring devices that can very accurately determine RR in a non-invasive way, to be used on a daily basis.

Regarding V_T measurement, current clinical methods include spirometry, impedance pneumography, inductance plethysmography, photoplethysmography, computed tomography, phonospirometry, Doppler radar, and more recently electrocardiography [9]–[17]. Similar to RR estimation, limitations arise when estimating V_T , e.g. high doses of ionizing radiation in computed tomography, or alteration in both natural RR and V_T due to spirometer use [18]. Moreover, having been designed for clinical settings or research centers, these methods employ specialized devices that are not translated easily to everyday use due to their high costs, need for skilled operators, or limited mobility.

Nowadays, smartphones are widely available and vital sign applications have been found to be accurate and robust. In

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

B. A. Reyes and N. Reljin are with the Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: bareyes@engr.uconn.edu; reljin@engr.uconn.edu).

Y. Kong and Y. Nam are with the Department of Computer Science and Engineering, Soonchunhyang University, Asan 336-745 Korea (e-mail: vengeful224@gmail.com; yynams@gmail.com).

*K. H. Chon is with the Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: kchon@engr.uconn.edu).

addition, smartphones have fast microprocessors, large data storage and media capabilities which make them an enticing option for developing a ubiquitous mobile respiration monitoring system. In an attempt to develop such a mobile system, we analyzed an acoustical approach and found good correlation between the smartphone-based respiration 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 a breathing range from 15 to 35 bpm [19]. However, the last approach requires plugging an additional acoustical sensor into the smartphone in order to extract information from tracheal sounds and just provides estimates of RR and breath-phase onset.

In order to overcome the need for an external sensor for the task of RR estimation, *i.e.*, the acoustical sensor, more recently our research group studied a noncontact optical approach that takes advantage of a smartphone's cameras. In particular, an algorithm that allows the real-time acquisition of a surrogate volumetric signal from breathing-related light intensity changes due to chest wall movements was implemented on a smartphone and its performance was tested in healthy volunteers breathing at a metered pace and spontaneously, while seated. Under this paced breathing, we found that the smartphone-based estimates of average RR were accurate when compared to those obtained from inductance plethysmography.

In general, the noncontact optical breathing monitor employs a video camera placed at distance from the subject's body to capture the intensity changes of the reflected light caused by his/her chest wall movements as they modify the path length of the illumination light [20]. During inspiration, the inspiratory muscles contract, resulting in an enlarged thoracic cavity; the diaphragm descends downward increasing the vertical dimension while the external intercostal muscles elevate the ribs and move the sternum upward and outward increasing the thoracic cavity in the horizontal axis. Due to this contraction the lungs expand to fill the larger thoracic cavity, resulting in a drop of the intra-alveolar pressure that causes a flow of air into the lungs until the intra-alveolar pressure equals the atmospheric pressure [21]. The inspiratory muscles relax during the expiration, restoring the chest wall and stretched lungs to their preinspiratory sizes, due to their elastic properties, and causing a rise in the intra-alveolar pressure above atmospheric level forcing the air to leave the lungs [21]. Note that in the noncontact optical respiratory monitoring approach, volume changes are not directly measured but a surrogate signal is obtained from the analysis of the variations in the reflected light due to chest wall movements captured by the system's camera while breathing.

There have been efforts to perform respiratory monitoring via the noncontact optical approach described above, but most of them have solely focused on average RR estimation [20], [22]–[27]. Still, noncontact optical methods have been proposed for V_T estimation, which is more challenging than average RR estimation. In particular, chest wall surface markers tracked by an optical reflectance system have shown promising results [23]. Those findings have been supported by studies that showed a one-to-one relationship between changes

of the external torso and V_T corresponding to internal lung air content [13]. More recently, a webcam and image processing technique based on the detection of shoulder displacements were implemented for breathing pattern tracking [25]. However, to the best of our knowledge, a smartphone-based system that uses a noncontact optical approach together with an algorithm implemented directly in the smartphone for the task of V_T estimation is not available yet.

Further observation of the smartphone-acquired signals during our previous study pointed us to the possibility of obtaining more valuable information than the average RR. Namely, we noticed that our algorithm was capable of monitoring the increased amplitude of the chest movements when volunteers took deeper breaths.

In this paper we propose a mobile system based on a noncontact optical approach implemented in a smartphone that provides information, from a volume surrogate, about both RR at each time instant (IRR) as well as V_T (when calibrated), in contrast to just average RR. For this study, the proposed respiratory monitoring system was implemented on a commercially-available Android smartphone, but could of course be implemented in smartphones using other operating systems. We collected signals from healthy volunteers and tested the performance of the proposed smartphone system for the tasks of IRR and V_T estimation, using the spirometer-acquired volume signal as reference.

II. MATERIALS AND METHODS

A. Subjects

For this study, fifteen ($N=15$) healthy and non-smoker volunteers (fourteen males and one female) aged 19 to 52 years (mean $\pm$ standard deviation: 28.73 ± 9.27), weight 70.14 ± 19.83 kg and height 175.67 ± 5.94 cm, were recruited. Exclusion criteria included individuals with previous pneumothorax, those with chronic respiratory illnesses such as asthma, and anyone who was currently ill with the common cold or an upper respiratory infection. The group of volunteers consisted of students and staff members from the University of Connecticut (UConn), USA. Each volunteer consented to be a subject and signed the study protocol approved by the Institutional Review Board of UConn.

B. Respiration signals acquisition

Equipment. The HTC One M8 smartphone (HTC Corporation, New Taipei City, Taiwan) running the Android v4.4.2 (KitKat) operating system was selected for this research as it is one of the state-of-the-art Android smartphones which is nowadays the dominant operating system worldwide in mobile devices. The HTC One M8 allows simultaneous dual camera recording supported by its processor running a 2.3 GHz quad-core CPU (Snapdragon 801, Qualcomm Technologies Inc., San Diego, CA, USA). For this study, the chest movement signal of interest was collected via the frontal camera consisting of a 5 MP, backside-illumination sensor with wide angle lens and 1080p full HD video recording capabilities at 30 frames-per-second. The video recording was processed in real time using

an application specifically designed for and implemented in the smartphone to obtain a volumetric surrogate signal, referred to in this paper as the chest movement signal, of the subject as discussed in the next section. After finishing the maneuver, the chest movement signal and corresponding time vector were saved into a text file in the smartphone and transferred to a personal computer for offline analysis of results using Matlab (R2012a, The Mathworks, Inc., Natick, MA, USA).

Together with the smartphone-recorded volumetric surrogate signal, a spirometer system consisting of a respiration flow head connected to a differential pressure transducer to measure airflow was used to record the airflow signal (MLT1000L, FE141 Spirometer, ADInstruments, Inc., Dunedin, New Zealand). The volume signal, regarded as reference for V_T and IRR estimation, was computed in the phone as the integral of the airflow over time. Both the airflow and volume signals were sampled at 1 kHz using a 16-bit A/D converter (PowerLab/4SP, ADInstruments, Inc., Dunedin, New Zealand). A 3.0 L calibration syringe (Hans Rudolph, Inc., Shawnee, KS, USA) was used to calibrate the spirometer system prior to recording each volunteer. A new set consisting of disposable filter, reusable mouthpiece, and disposable nose clip was given to each volunteer (MLA304, MLA1026, MLA1008, ADInstruments, Inc., Dunedin, New Zealand).

Acquisition protocol. Each maneuver lasted approximately 2 minutes during which the volunteers were asked to breathe through the spirometer system at different volume levels ranging from around 300 mL to 3 L depending on what manageable for that individual. Each subject was instructed to breathe while first increasing their V_T with each breath for

around 1 minute, and then decreasing their V_T with each breath for the remaining time. To provide visual feedback of the maneuver to the volunteers, their volume signal was displayed on a 40" monitor placed in front of them. Nose clips were used to clamp the nostrils during the respiration maneuver. Subjects were standing still during signal collection. The smartphone was positioned in front of the subject at approximately 60 cm in a 3-pronged clamp placed at thorax level so that the frontal camera recorded chest wall movements associated with breathing during the maneuver. All signals were recorded in a regular dry lab with the ambient light which predominantly consisted of ordinary fluorescent lamps located in the ceiling approximately 2.5 m above floor level and to a lesser extent, sunlight entering through the lab's windows. Although the smartphone and spirometer recordings were simultaneously started, 5 seconds of initial and final apnea segments were acquired for automatic alignment purposes between both recordings. After initial apnea, subjects took a forced respiration cycle before performing the described respiration maneuver. Fig. 1 shows an example of the experimental setup. It is worth mentioning that volunteers were not restricted in wearing any color/pattern of their clothes during the maneuvers but instructed not to wear loose clothes.

C. Chest movement recording algorithm

The two major anatomical contributors to the visibility of breathing are the rib cage and abdomen compartments of the chest wall, whose movements in the anteroposterior direction are greater than those in the vertical or transverse directions, with an increase of around 3 cm in the anteroposterior diameter

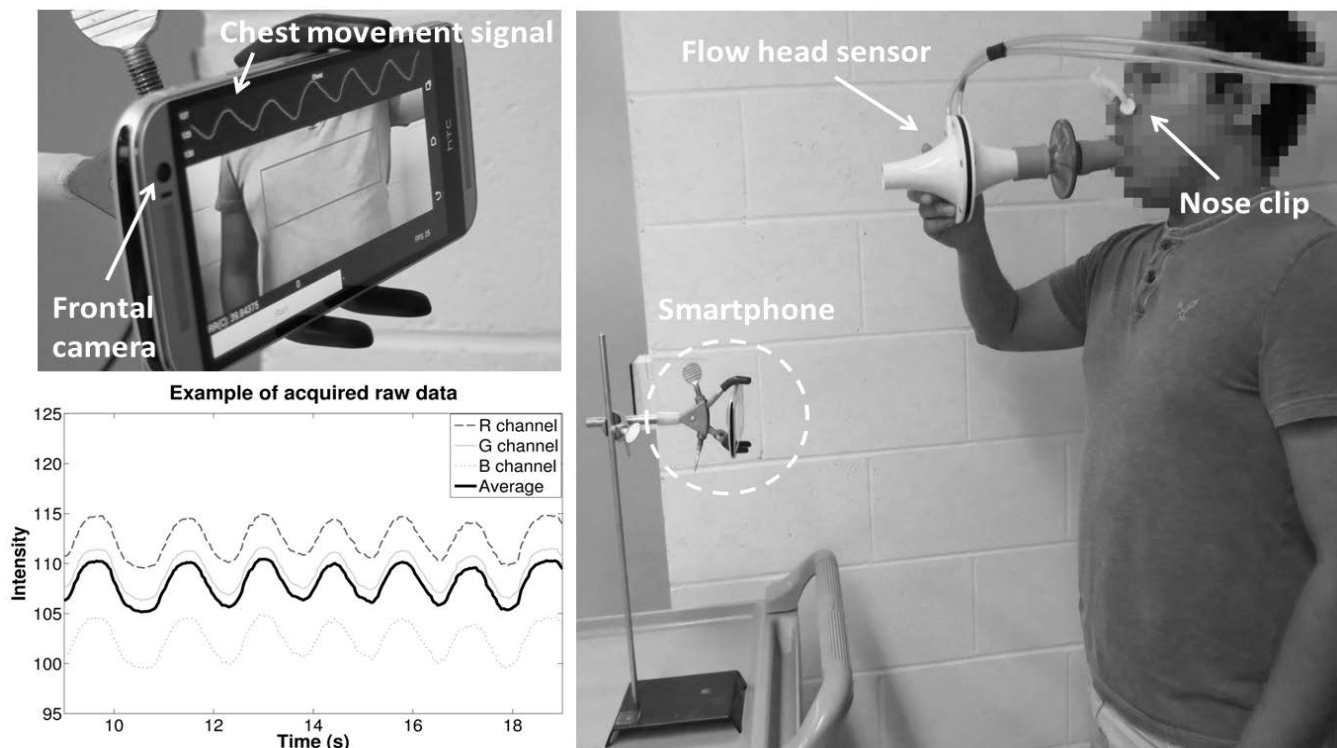


Fig. 1. Recording of chest movement signal using a smartphone's camera and volume using a spirometer during a respiration maneuver. The smartphone was placed in front of the subject at thorax level in order to record chest movements which were later compared to the reference volume signal from the spirometer. *Top left panel:* Zoomed view of the developed smartphone app. *Bottom left panel:* A segment of the raw signals extracted from the RGB channels and their average.

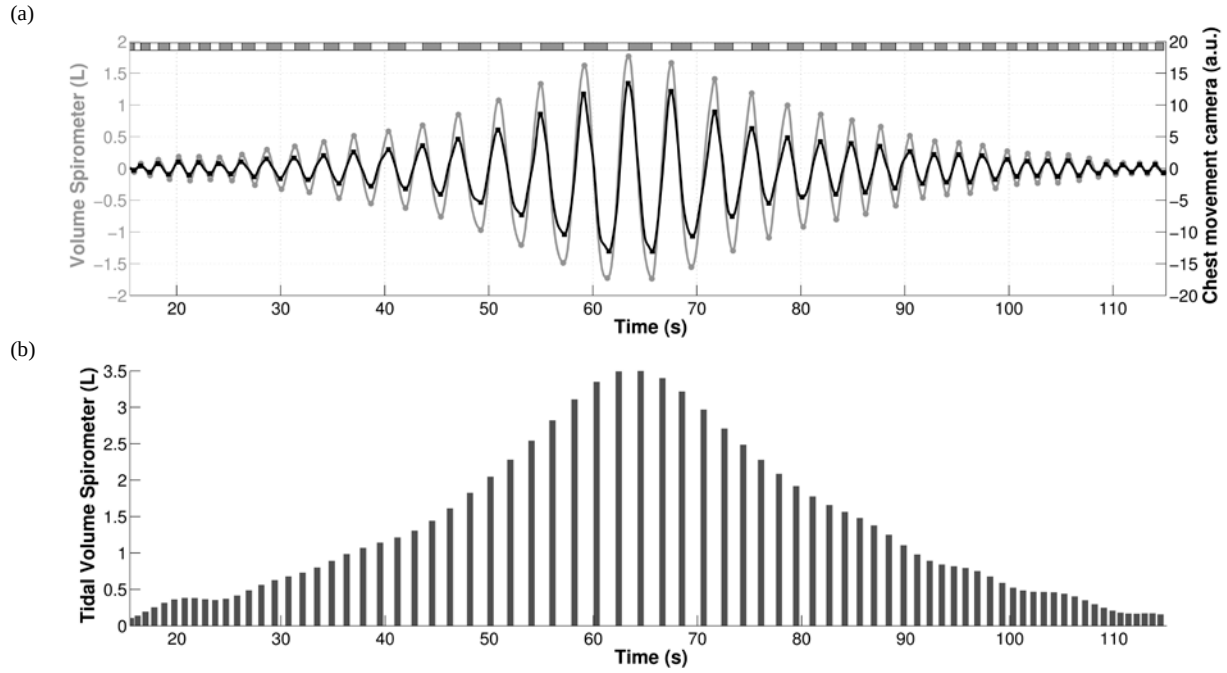


Fig. 3. Example of pre-processed signals during the respiration maneuver of one subject. (a) Detrended versions of a volume signal from the spirometer and the chest movement signal from a smartphone camera. Gray and black dots indicate the maxima and minima of volume and chest movement signals,¹ respectively. (b) Tidal volume of each respiration phase computed as the absolute difference between the volumes at two consecutive breath-phase onsets.^t

over the vital capacity range [28]. There is a relationship between volume displacement and linear motion during breathing [28], and a one-to-one relationship between changes of the external torso and tidal volume corresponding to internal lung air content has been found [13]. The proposed smartphone algorithm is intended to take advantage of this relationship to obtain a volumetric surrogate by analyzing the changes in the intensity of the reflected light caused by the breathing-related chest wall movements captured at a distance with a smartphone's camera. In particular, the algorithm processes video recordings in real time, where at each time instant t , the intensities of the red, green and blue (RGB) channels are averaged within a rectangular region of interest (ROI) according to

$$I(t) = \left(\frac{1}{3D} \right) \left(\sum_{\{m,n\} \in ROI} i_R(m,n,t) + \sum_{\{m,n\} \in ROI} i_G(m,n,t) + \sum_{\{m,n\} \in ROI} i_B(m,n,t) \right) \quad (1)$$

where $i_x(m,n,t)$ is the intensity value of the pixel at the m -th row and n -th column of the red, green or blue channel within the ROI containing a total of D pixels. For this study, a region of 49×90 pixels were selected in a resolution of 320×240 pixels and focused on the thoracic area of the subject. This reduced resolution and ROI size were selected so that they do not compromise the sampling rate during the real time monitoring in the smartphone app. With these settings, the frame rate

dropped to around 25 frames-per-second. The average intensity waveform $I(t)$ was regarded as the chest movement signal, *i.e.*, the volume surrogate, from which the tidal volume and respiratory rates were estimated. As shown in Fig. 1, despite the DC values all channels carry similar information, and hence their average was taken to avoid channel selection. An example of the raw volume acquired with a spirometer and the corresponding chest movement signal acquired online with the smartphone's camera and chest movement app is shown in Fig. 2 for the respiration maneuver performed by one subject. It should be noted that similar to other monitoring methods, *e.g.* inductance plethysmography, the proposed noncontact optical approach via the smartphone-acquired volumetric surrogate signal might be very weak if the clothes worn by the subject are not tight to his/her thorax, which can result in increased estimation errors of breathing parameters.

D. Data preprocessing

The acquired chest movement signal was interpolated at 25 Hz via a cubic spline algorithm to achieve a uniform sampling rate that corrects fluctuations around this value during the online acquisition in the smartphone. The reference volume signal was down-sampled to 25 Hz to achieve the same sampling frequency as the chest movement signal. In order to minimize high frequency components not related to the respiration maneuver, the chest movement and reference volume signals were filtered with a 4th-order Butterworth lowpass filter at 2 Hz that was applied in a forward and backward scheme to produce zero-phase distortion and minimize the start and end transients.

After filtering, the chest movement and reference volume signals were automatically aligned using the cross-correlation

function, where 20 seconds in the central portion of the maneuver were extracted from each recording to compute the cross-correlation sequence in order to obtain the sample lag providing the maximum cross-correlation value that indicates the required samples to be shifted. This alignment was required because of different starting times and delays of the smartphone and AD converter acquisition systems during the simultaneous recording of the maneuver. The duration of the signals was set accordingly, to the minimum duration of both types of recordings.

Finally, both signals, the surrogate and actual volume, were detrended via the Empirical Mode Decomposition (EMD) method [29]. The essence of this decomposition is to identify the intrinsic oscillatory modes, called IMFs, of a signal through the time scales present in it. Its principal attractiveness resides in obtaining the IMFs directly from the signal without the use of any kernel, i.e., EMD depends only on the data. All the IMFs of the signal $s(t)$ under analysis are extracted automatically by a shifting process intended to eliminate riding waveforms and to produce close to zero mean value as defined by upper and lower envelope signals. The EMD sifting process allows representation of the original signal in term of its extracted components as

$$s(t) = \sum_{k=1}^K IMF_k(t) + r_K(t) \quad (2)$$

where K is the total number of IMFs, and $r_K(t)$ is the residual signal. EMD has the characteristic of being a complete decomposition [29].

E. Tidal volume estimation using smartphone camera signal

The volume signal from the spirometer was used to automatically determine the breath-phase onsets during the maneuver by finding their local maxima and minima. Inspiratory and expiratory phases corresponded to positive and negative traces of the volume signal, respectively. The V_T of each phase was computed as the absolute volume difference between two consecutive breath-phase onsets. The time location of the onsets was used to determine the corresponding maxima or minima in the aligned chest movement signal around a time window of 500 ms centered at each breath-phase onset. The amplitude difference between two consecutive breath-phase onsets in the chest movement signal was used for V_T estimation via the smartphone.

For calibration, a least-squares linear regression between the reference V_T and the absolute peak-to-peak amplitude of chest movement was performed for each subject; half of the data points of the maneuver were randomly selected for calibration purposes and regarded as a training data set, while the remaining half were used as a test data set to which the computed linear model was applied in order to map the smartphone-based measurements to volume estimates in liters.

The performance of the V_T estimation was measured on the test data using the regression parameter r^2 , the root-mean-squared error $RMSE$, and the normalized root-mean-squared error $NRMSE$, defined as follows

$$RMSE = \sqrt{\frac{\sum_{i=1}^M (V_{T_{spirometer}}(i) - V_{T_{smartphone}}(i))^2}{M}} \quad (3)$$

$$NRMSE = \frac{RMSE}{\text{mean}(V_{T_{spirometer}})} \times 100\% \quad (4)$$

where $V_{T_{spirometer}}$ indicates the tidal volume obtained from the spirometer-acquired volume signal, $V_{T_{smartphone}}$ the tidal volume estimated from smartphone-acquired chest movements after calibration, and M is the number of breath-phases of the analyzed maneuver used for testing.

Fig. 3 shows an example of the preprocessed reference volume and chest movement signals. The breath-phase onsets and respiration phases as computed from the volume signal are indicated on top. The corresponding maxima and minima are superimposed on each signal. Detrended versions of the signals are presented. The corresponding V_T of each respiration phase, computed as the absolute volume difference between two consecutive breathing onsets, is also shown below the respiration maneuver.

F. Instantaneous respiration rate estimation using smartphone camera signal

To estimate IRR from the smartphone-acquired chest movement signal, a time-varying spectral technique was used. In this paper, the smoothed pseudo Wigner-Ville distribution (SPWVD) time-frequency representation (TFR) was employed. A TFR is a function that simultaneously describes the energy density of a signal in the time and frequency domains, allowing one to analyze which frequencies of a signal under study are present at a certain time [30]. TFR analysis is useful for analyzing signals whose frequency content varies over time, as is the case with respiration signals.

The Wigner-Ville distribution (WVD) belongs to the Cohen's class of bilinear time-frequency representations; it possesses several interesting properties, and in particular provides the highest time-frequency resolution. However, the main limitation of the WVD is the presence of cross-terms that obscure its readability. Several techniques have been proposed to reduce the number of cross-terms of the WVD; however, there is a tradeoff between the amount of cross-term interference and the time-frequency resolution. The spectrogram is one such attempt, a joint time-frequency smoothing window is applied and hence the performance in one direction is enhanced at the expense of degrading the performance in the other. In contrast, the SPWVD employs independent time and frequency smoothing windows [31], as given by

$$SPWVD(t, f) = \int_{-\infty}^{\infty} h(\tau) \int_{-\infty}^{\infty} g(\eta - t) \cdot s\left(\eta + \frac{\tau}{2}\right) s^*\left(\eta + \frac{\tau}{2}\right) d\eta e^{-j2\pi f\tau} d\tau \quad (5)$$

where $s(t)$ is the signal under analysis, $g(\cdot)$ is the time smoothing window, and $h(\cdot)$ is the frequency smoothing

The SPWVD was applied to the volume and chest movement signals. The SPWVD was computed using $NFFT=1024$ frequency bins, a 2 second Hamming window as the time smoothing window, and a 5.12 second Hamming window as the frequency smoothing window. After computing, the SPWVD was normalized between [0-1]. The Welch modified periodogram was used to compute the spectrum of the whole maneuver in order to obtain the central or average respiration frequency as the maximum spectral peak. The periodogram was computed using 50% overlap, 512 frequency bins, and a Hamming window. Then, at each time instant the maximum peak around the central frequency was computed and the frequency at which that maximum occurs was regarded as the respiration frequency at that instant, so that a vector of instantaneous respiration frequency was returned from each SPWVD. The frequency vector extracted from the spirometer-based volume was regarded as the reference instantaneous respiration frequency and was compared against the frequency vector extracted from the corresponding smartphone-based chest movement signal. All instantaneous respiration frequencies were converted from hertz to breaths-per-minute (bpm) to obtain IRR.

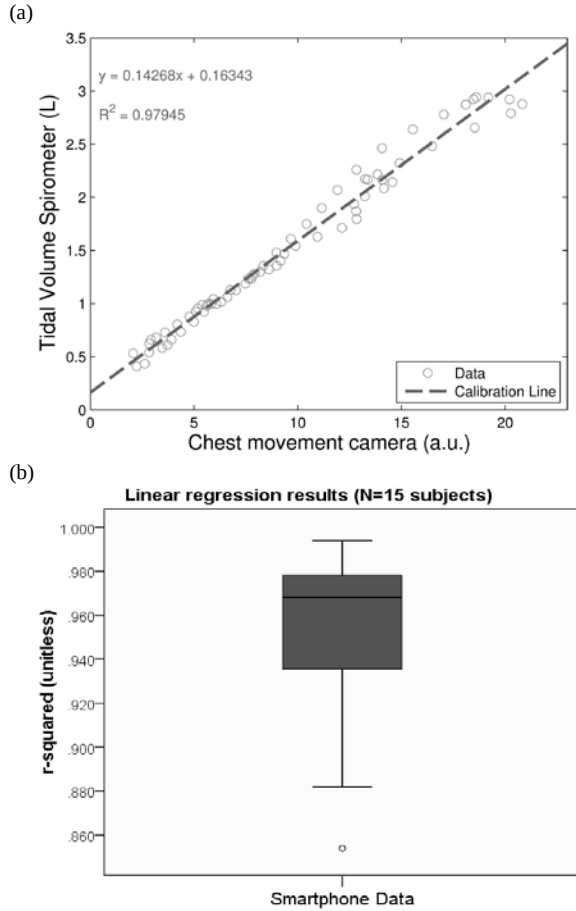


Fig. 4. Least-squares linear regression between the chest movement amplitude differences from smartphone and reference tidal volume from spirometer. (a) Example of regression for all data from one subject. (b) Boxplot of r^2 regression parameter for all subjects.

window in the time-domain [32].

Similar to tidal volume estimation, the performance of the IRR estimation using the smartphone-acquired chest movement signal was tested using three performance indices by considering the IRR from volume signal as reference: the root-mean-squared error $RMSE$, the normalized root-mean-squared error $NRMSE$, and the cross-correlation index ρ defined as follows

$$\rho = \frac{\sum_{i=1}^S IRR_{spirometer}(i) \cdot IRR_{smartphone}(i)}{\sqrt{\sum_{i=1}^S (IRR_{spirometer}(i))^2 \cdot \sum_{i=1}^S (IRR_{smartphone}(i))^2}} \quad (6)$$

where $IRR_{spirometer}$ indicates the IRR obtained from the spirometer-acquired volume signal, $IRR_{smartphone}$ is the IRR estimated from smartphone-acquired chest movements, and S is the number of samples of the analyzed signal, i.e., time instants. $RMSE$ and $NRMSE$ were computed via (3) and (4), by replacing the V_T values at each breath-phase by the IRR values at each time instant.

III. RESULTS

The smartphone-acquired chest movement signal showed temporal amplitude variation related to the volume from spirometer during the breathing maneuver as shown in Fig. 2 and more evidently in Fig. 3 after detrending. In the following subsections we present the results in terms of tidal volume estimation and respiration rate estimation using this smartphone-acquired chest movement signal. The distribution of the number of breathing cycles, average V_T , and average RR performed by volunteers during the breathing maneuvers are shown in Table I. As can be seen, the maneuvers included a wide range of breathing cycles, rates and depths.

A. Tidal volume estimation using smartphone camera signal

Fig. 4 shows the relationship between the absolute peak-to-peak amplitude of chest movement acquired with the smartphone and the reference tidal volume acquired with the spirometer for each breath phase of the maneuver performed by one subject. As shown in this figure, the amplitude differences of smartphone-based chest movement signals linearly correlate to reference V_T from the spirometer. The regression parameter r^2 between the absolute peak-to-peak amplitude of chest movement and reference tidal volume was computed for all breath-phases of each subject ($r^2=0.951 \pm 0.042$, mean $\pm$ SD). The corresponding boxplot for all subjects is also shown in Fig. 4. Strong linear relationship ($r^2>0.9$) was found between the smartphone-based estimates and the reference tidal volume from the spirometer, as tested via a one-sample Wilcoxon signed rank test ($p=6.41 \times 10^{-4}$) after the normality assumption did not hold (one-sample Kolmogorov-Smirnov test, $p=0.002$).

An example of the V_T estimation procedure from smartphone-acquired data is shown in Fig. 5. From top to bottom, the first two plots of Fig. 5 correspond to the calibration process using the training data set (Fig. 5a), and the testing process using the remaining randomly-selected breath-phase data points (Fig. 5b), respectively. The calibration parameters were computed via least-squares linear regression. Fig. 5c

shows the corresponding smartphone-based V_T estimates, after using the calibration parameters, for each breath phase of the maneuver of one subject. The lower panel of Fig. 5c shows the corresponding error differences with respect to the reference V_T

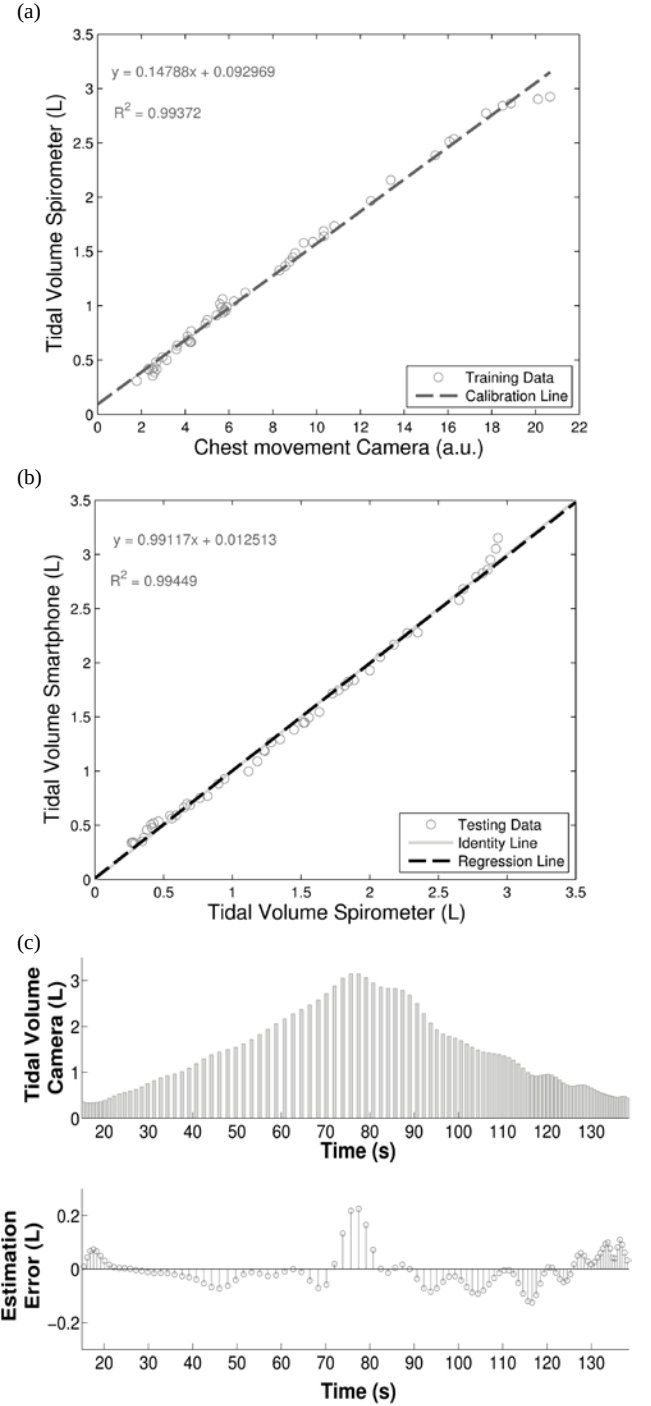


Fig. 5. Example of tidal volume estimation using a smartphone for one subject. (a) Calibration curve between chest movement amplitude differences from smartphone and tidal volume from spirometer for half of data randomly selected as training data. (b) Linear regression between smartphone-estimated tidal volume, after applying the calibration linear model, and reference tidal volume for the remaining half. (c) *Top*: Estimated tidal volumes from smartphone for each breath phase of the maneuver. *Bottom*: Corresponding differences between tidal volume derived from smartphone and reference tidal volume from spirometer throughout the whole breathing maneuver.

TABLE I.
DISTRIBUTION OF BREATHING CYCLES, TIDAL VOLUME AND RESPIRATION RATE MEASURED BY SPIROMETER
DURING BREATHING MANEUVERS (N=15 SUBJECTS).

Parameter		Min	Max	Average
Breathing cycles	[cycles]	16	51	31.40 ± 10.25
Maneuver tidal volume	[L]	0.24 ± 0.11	3.11 ± 0.67	1.32 ± 0.26
Maneuver respiration rate	[bpm]	11.08 ± 3.69	35.45 ± 13.04	17.12 ± 5.28

Values presented as mean ± standard deviation

from spirometry.

The performance indices for smartphone-based V_T estimation are presented in Table II for the testing data set of all the volunteers, using the spirometer measurements as reference. The linear regression results shown in Fig. 5, for one subject, hold for all subjects, as shown in Fig. 6, when a linear regression was applied to all the tidal volume estimates from all volunteers. Fig. 6 also presents the corresponding Bland-Altman plot.

We found that when calibrated on a subject-by-subject basis, the smartphone-based V_T estimation produced a bias of 0.014 liters and a standard deviation of 0.185 liters, however the bias was not found to be statistically significant from a zero bias. Accordingly, the 95% limits of agreements were -0.348 to 0.376 liters.

B. Instantaneous respiration rate estimation using smartphone camera signal

Fig. 7 shows an example of IRR estimation via the SPWVD technique applied to volume from a spirometer and chest movements from the smartphone for the respiration maneuver of one subject. The superimposed white dashed curve indicates the frequency at which the maximum energy of the SPWVD occurs at each time instant. Side-by-side comparison of the extracted IRR from spirometer and smartphone signals is also presented.

Table III presents the performance indices of smartphone-based IRR estimation for all the subjects, using the spirometer values as reference. High cross-correlation coefficients were found between the IRR smartphone-based estimates and volume from spirometer. Fig. 8 reflects this high correlation as shown by the regression line parameters ($r^2=0.9973$). The corresponding Bland-Altman plot is also presented in Fig. 8. Compared to the spirometer, the bias ± standard deviation and the 95% limits of agreement were -0.024 ± 0.421 bpm and -0.850 to 0.802 bpm, respectively. Note that in this Bland-Altman plot, the IRR differences distribute at regular intervals given by the width of the frequency bins used in the calculation of the FFT during the time-frequency analysis, $\Delta = \frac{fs/2}{NFFT} = 0.0122$ Hz equivalent to $\Delta = 0.7324$ bpm.

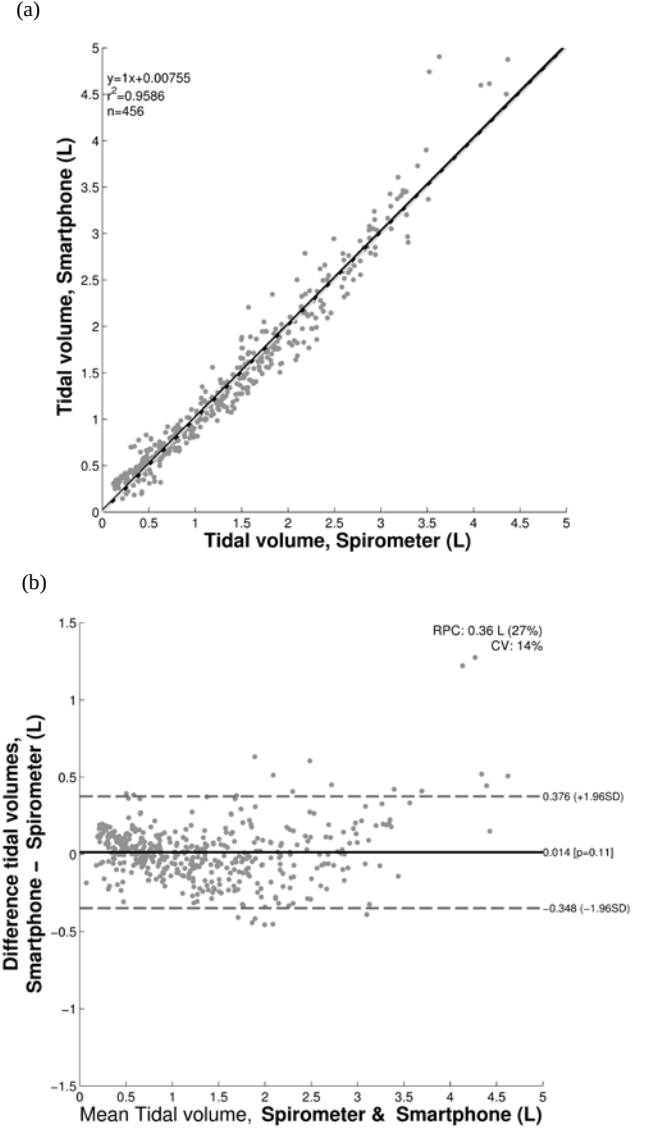


Fig. 6. Tidal volume estimation using smartphone (N=15 subjects). (a) Regression curve. Gray dashed line indicates the identity line and the solid black the regression line. (b) Bland-Altman plot. Solid black line indicates the bias and dashed gray lines indicate the 95% limits of agreement.

TABLE II.
RESULTS OF TIDAL VOLUME ESTIMATION USING SMARTPHONE-ACQUIRED CHEST
MOVEMENT SIGNALS COMPARED TO THE REFERENCE VOLUME
FROM THE SPIROMETER (N=15 SUBJECTS).

Parameter		Values		
r^2	[unitless]	0.961	$\pm$	0.026
RMSE	[L]	0.182	$\pm$	0.107
NRMSE	[%]	14.998	$\pm$	5.171

Values presented as mean $\pm$ standard deviation

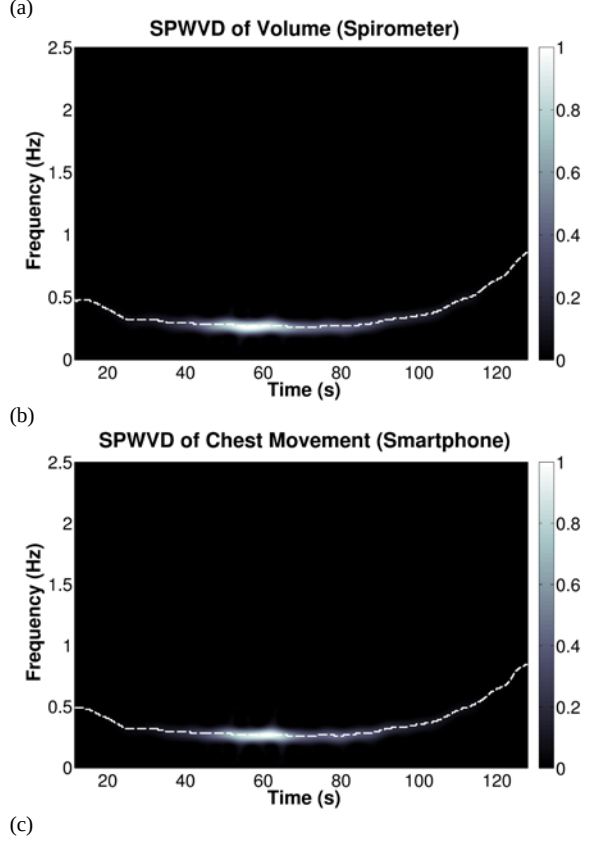


Fig. 7. Example of IRR estimation using smartphone-acquired chest movement signals. (a) SPWVD of the volume signal from the spirometer. (b) SPWVD of the chest movement signal from smartphone. White dashed lines indicate the maximum peak at each time instant. (c) Instantaneous respiration rate computed from corresponding SPWVD of volume and chest movement signals.

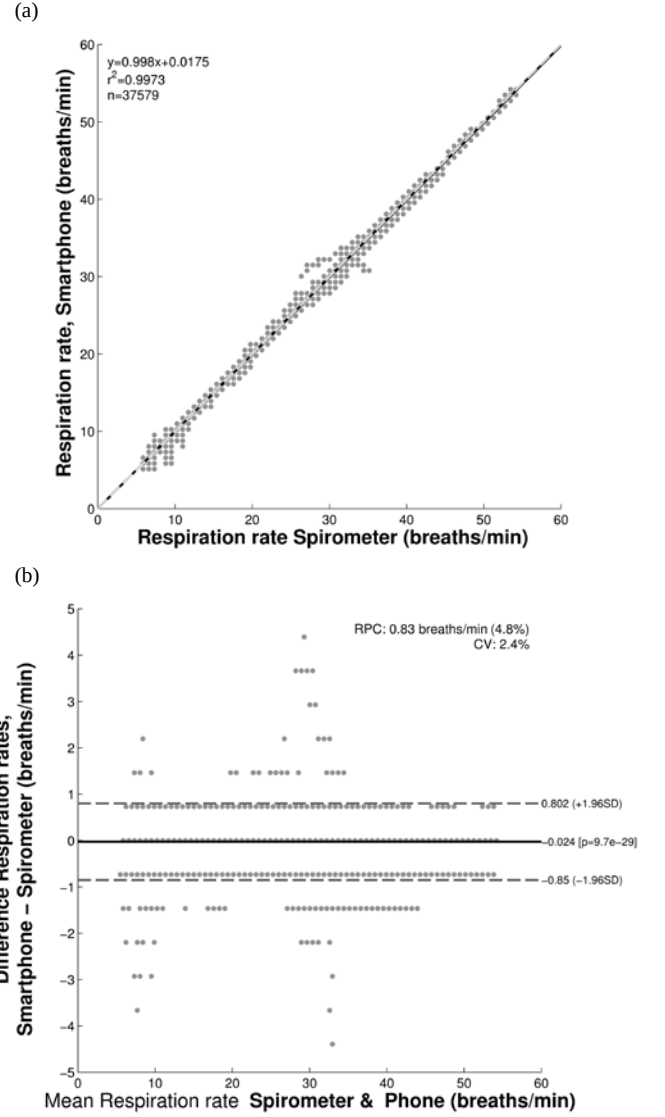


Fig. 8. Instantaneous respiration rate estimation using smartphone (N=15 subjects). (a) Calibration curve. Gray dashed line indicates the identity line and the solid black the calibration line. (b) Bland-Altman plot. Solid black line indicates the bias and dashed gray lines indicate the 95% limits of agreement.

TABLE III.
RESULTS OF THE INSTANTANEOUS RESPIRATION RATE ESTIMATION USING
SMARTPHONE-ACQUIRED CHEST MOVEMENT SIGNAL COMPARED TO
VOLUME SIGNAL FROM SPIROMETER (N=15 SUBJECTS).

Parameter		Values		
ρ	[unitless]	0.9992	$\pm$	0.0019
RMSE	[bpm]	0.414	$\pm$	0.178
NRMSE	[%]	3.031	$\pm$	2.873

Values presented as mean $\pm$ standard deviation

IV. DISCUSSION AND CONCLUSIONS

In this paper we propose a smartphone-based respiration monitoring system for both instantaneous respiration rate estimation and tidal volume estimation via an algorithm that tracks chest movement directly from a smartphone's camera. The HTC One M8 Android smartphone was used in this study and the algorithm was implemented in this device so that recordings of the chest movement signals were made directly on the phone. Together with this smartphone signal, airflow and volume signals were recorded with a spirometer and the latter was used as reference for IRR and V_T estimation. Recordings from fifteen healthy volunteers were obtained in a regular dry lab illuminated with fluorescent light while the volunteers were standing still and breathing at tidal volumes ranging from 300 mL to 3 L. Volunteers wore clothes with different colors and patterns. The developed algorithm can still detect the chest movements even if single color clothes are worn.

There have been several efforts to develop monitors that provide information about breathing status via optical approaches [20], [22]–[27], most of them monitoring only average RR. Recently, we implemented an algorithm that is able to track chest movements directly on a smartphone and found promising results in terms of average RR estimation. That study provided motivation to explore whether information beyond the average RR can be obtained from the smartphone-acquired chest movement signal. In particular, it appeared that the smartphone app provided a signal whose peak-to-peak amplitude was an indicator of the tidal volume of the volunteers. This hypothesis was corroborated in this study as exemplified in the recorded reference volume and chest movement signals, especially after detrending via EMD to remove existing drift in both signals. We also analyzed the correlation of the peak-to-peak amplitude of smartphone-acquired signals with the corresponding tidal volume signal acquired from a spirometer. We found that a strong correlation existed between the peak-to-peak amplitude of chest movement signals and tidal volume from the spirometer ($r^2=0.951 \pm 0.042$, mean $\pm$ SD). Given these correlation results, for each subject we randomly selected 50% of the data points for training the linear model during the calibration process, and the remaining 50% of the data for testing the tidal volume estimation based on the computed model. Once calibrated on an individual basis using the reference volume signal, when we mapped the chest movement amplitude differences at each breath-phase of the testing data set, we found an *RMSE* of 0.182 ± 0.107 liters which corresponded to 14.998 ± 5.171 % when normalized to the mean value of the reference V_T of the testing data set of the maneuver. Overall, we found that a linear regression model fitted well the calibrated peak-to-peak amplitude of smartphone signals for the task of V_T estimation ($V_{T\text{smartphone}}=1.005 \cdot V_{T\text{spirometer}}+0.008$). We did not find statistically-significant bias in the V_T estimation using smartphones and the 95% limits of agreement were -0.348 to 0.376 liters. At this point it is difficult to state if this error estimate in tidal volume is acceptable for home monitoring use. Other popular methods for tidal volume estimation suffer from

even higher estimation errors, for example, respiratory inductance plethysmography (RIP), when calibrated according to the manual (which usually states that 10% error difference is acceptable), often has much higher errors. Others have reported similar findings with respect to errors, e.g., reference [5] found a bias and 95% limits of agreement in RIP sensors of approximately 0.4 L, and -0.3 to 1.1 L for a breathing range of 360 mL to 3.5 L. As noted in this work, the estimation error using RIP is even higher than our proposed approach using a smartphone's video camera.

By taking advantage of the high correlation between detrended smartphone signals and volume from the spirometer, we analyzed using the smartphone signal for the task of RR estimation at each time instant. Due to the time-varying characteristics of the signals, we employed the smoothed pseudo Wigner-Ville distribution. We found high correlation between the smartphone-based IRR estimates and the spirometer-based values ($r^2=0.9992 \pm 0.0019$). We found an *RMSE* of 0.414 ± 0.178 bpm which corresponds to an *NRMSE* of 3.031 ± 2.783 %. The linear relationship between IRR estimated from the smartphone and IRR from reference volume was $IRR_{\text{smartphone}}=0.9980 \cdot IRR_{\text{spirometer}}+0.0175$. The 95% limits of agreement ranged from -0.850 to 0.802 bpm, while there was a statistically-significant bias of -0.024 bpm. Other studies have reported the estimation of respiratory rate using noncontact optical approaches, e.g., in [22] the bias and standard deviation were found to be 0.19 bpm and 2.46 bpm, respectively, in the range of approximately 10-70 bpm; in [24] the *RMSE*, bias, and standard deviation were 1.28 bpm, 0.12 bpm, and 1.33 bpm, respectively, in the range of approximately 10-22 bpm; in [25] the *RMSE*, bias and 95% limits of agreement were 1.20 bpm, 0.02 bpm, -2.40 to 2.45 bpm, respectively, in the range of approximately 10-24 bpm; while in [20] the *RMSE*, bias, and 95% limits of agreement were found to be 0.09 bpm, -0.02 bpm, and -1.69 to 1.65 bpm, respectively, in the range of approximately 7-24 bpm. Interestingly, the results reported in [20] during night conditions outperformed those mentioned in the several sentence above during daylight conditions. Although a straightforward comparison is not possible due to the differences in the measurement devices and the noncontact distance ranges tested, in general, the current proposed noncontact optical monitoring of respiratory rate based on smartphones performs as well as, if not better than, the aforementioned studies.

Limitations of this study include the recording of the breathing maneuvers while the subjects were standing still, i.e., the subjects were instructed not to move. As found in other noncontact optical approaches, the main challenge arises from motion artifacts, especially when the dynamics of both the volumetric surrogate signal obtained from the chest wall movements and the motion artifacts have similar low frequency ranges (<2 Hz). Hence, it is expected that motion artifacts deteriorate the performance of the smartphone-based breathing estimates. Implementation of body tracking and artifact removal schemes similar to those reported in the literature to improve respiratory rate estimation [25], [33] are expected to reduce the effect of body motion not related to the breathing

maneuver. Implementation and testing of such algorithms in the smartphone for respiratory monitoring, especially for the task of tidal volume estimation, should be explored in future studies.

Another major challenge is the variation of the ambient illumination at different times of the day due to fluctuations in the amount of sunlight, for example. The experiments presented in our study were performed at different times of the day and while the main illumination source came from the ceiling fluorescent lamps, the window shades of the laboratory were kept open or closed according to the needs of its users. Despite that, we did not notice that these variations disturbed the acquisition of the volumetric surrogate signal, perhaps due to the dominance of the fluorescent source. We recognize that systematic studies must analyze the performance of the proposed system in different levels of ambient illumination as well as to explore ways to account for these illumination variations.

The chest wall area of interest monitored during the breathing maneuver also represents another limitation. Classically, chest wall movements are attributed to two mechanical degrees of freedom due to contributions from rib cage and abdomen, which can be used to estimate tidal volume [28]. Although 1D or 2D displacements of these two compartments account for the majority of tidal volume, the algorithm ignores systematic effects of rib cage distortions [23]. In this study, the chest movement signal used as volume surrogate was extracted from an image's rectangular area centered on the anterior chest wall portion of the volunteer that visually provided the most dominant displacements while breathing. Accordingly, our approach ignores those small contributions due to rib cage distortions and only constructs the chest movement signal from the chest wall displacements monitored by the camera.

It is also expected that postural changes and airway obstruction impact the performance of the estimates, as has been found in other breathing monitor techniques [34], [35]. Postural changes can modify the contribution of the rib cage and abdomen compartments to tidal volume. A decreased rib cage excursion and an increased abdominal excursion have been found in the supine position compared to the sitting or standing postures [36], [37]. Accordingly, it is likely that another area of the thorax would provide a stronger surrogate signal when monitoring breathing in the supine position.

Another limitation is requiring subjects to wear fitted clothes during the experiments. As pointed out by other researchers, if the clothes are not tight enough to the subject's body a weak breathing-related signal might be obtained using the noncontact optical monitoring approach. Note that this is also the case in other respiratory monitoring methods based on chest wall displacements, like inductance plethysmography, where the sensors are recommended to be worn over bare skin or tight clothes. Observe that in general, the noncontact optical approach looks for changes in the light intensity due to the modification of the path length caused by breathing displacements of the chest wall, and is not limited to movements of clothing features. However, a systematic study is required to analyze the effect of wearing loose-fitting clothes.

Finally, at this research point, note that to estimate tidal volume via the smartphone's camera, the measurement conditions should match those during which calibration was performed. Although we found that a linear model fitted well between peak-to-peak amplitude of chest movement signals from a smartphone and tidal volume from a spirometer, so that it can be used to calibrate the smartphone measurements to obtain tidal volume on an individual basis, a new calibration should be performed prior to acquisition if the subject's chest wall position monitored by the smartphone's camera displaces with respect to the one used for calibration. Other tidal volume estimation techniques suffer similar issues, *e.g.*, displacement of elastic belts wrapped around the rib cage and abdomen from the position employed when calibration was performed deteriorates the performance of the measurements in inductance plethysmography.

Several monitoring techniques for breathing status in clinical and research settings currently exist. This study and similar works are steps towards the developing of an inexpensive and mobile respiratory monitoring system that can be translated outside research settings for on-demand health applications. By taking advantage of their ubiquity, smartphone-based systems could aid in the monitoring of breathing status of the general population, where this general practice remains unclear if we consider that these parameters are not always recorded on a daily basis even in clinical settings. The results obtained in this study point out the feasibility of developing a mobile system being able to provide information about instantaneous respiration rate and tidal volume when calibrated on an individual basis. It is foreseen that when calibration is not possible to be performed, this smartphone approach could still be used as a qualitative indicator of changes in tidal volume due to the high correlation between the chest movement signal and tidal volume that reflects the major contribution of chest wall displacements to tidal volume. To this end, this paper reports our initial step towards the estimation of V_T from a surrogate signal obtained with a smartphone. We cannot make conclusions about the robustness in terms of measurement conditions such as gender, body mass index or lighting conditions given the small sample size and conditions tested. These will be subjects to be explored in future studies.

Currently, we are running a study regarding an easy-to-use calibration procedure that can be performed with an incentive spirometer (IS) with the potential to be translated outside research settings due to their availability and low cost. Briefly, by taking advantage of the high linear relationship between smartphone measurements and tidal volume, the calibration model is computed while breathing at only two reference volume points through the IS. Preliminary results have shown to be comparable to those presented in this paper which would allow a fast and easy-to-perform calibration procedure. In parallel, we are currently working on the implementation of the proposed signal processing techniques, currently developed for Android, on iOS to cover the two dominant smartphone operating systems. We envision a subject using the proposed tool in their home as an alternative to a spirometer. The person would place their smartphone at a fixed location, stand still in front of it and conduct a series of breathing routines, and obtain

some measurements. By doing so, this would minimize the motion artifacts.

REFERENCES

- [1] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatr. Pulmonol.*, vol. 46, no. 6, pp. 523–529, Jun. 2011.
- [2] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign," *Med. J. Aust.*, vol. 188, no. 11, p. 657, 2008.
- [3] M. Folke, L. Cernerud, M. Ekström, and B. Hök, "Critical review of non-invasive respiratory monitoring in medical care," *Med. Biol. Eng. Comput.*, vol. 41, no. 4, pp. 377–383, Jul. 2003.
- [4] B. M. Koeppen and B. A. Stanton, *Berne & Levy Physiology, Updated Edition*. Elsevier Health Sciences, 2009.
- [5] K. P. Cohen, W. M. Ladd, D. M. Beams, W. S. Sheers, R. G. Radwin, W. J. Tompkins, and J. G. Webster, "Comparison of impedance and inductance ventilation sensors on adults during breathing, motion, and simulated airway obstruction," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, pp. 555–566, Jul. 1997.
- [6] G. B. Drummond, A. F. Nimmo, and R. A. Elton, "Thoracic impedance used for measuring chest wall movement in postoperative patients," *Br. J. Anaesth.*, vol. 77, no. 3, pp. 327–332, Sep. 1996.
- [7] M. A. E. Ramsay, M. Usman, E. Lagow, M. Mendoza, E. Untalan, and E. De Vol, "The Accuracy, Precision and Reliability of Measuring Ventilatory Rate and Detecting Ventilatory Pause by Rainbow Acoustic Monitoring and Capnometry," *Anesth. Analg.*, vol. 117, no. 1, pp. 69–75, Jul. 2013.
- [8] J. J. Vargo, G. Zuccaro Jr., J. A. Dumot, D. L. Conwell, J. B. Morrow, and S. S. Shay, "Automated graphic assessment of respiratory activity is superior to pulse oximetry and visual assessment for the detection of early respiratory depression during therapeutic upper endoscopy," *Gastrointest. Endosc.*, vol. 55, no. 7, pp. 826–831, Jun. 2002.
- [9] K. Ashutosh, R. Gilbert, J. H. Auchincloss, J. Erlebacher, and D. Peppi, "Impedance pneumograph and magnetometer methods for monitoring tidal volume," *J Appl Physiol*, vol. 37, no. 6, pp. 964–966, 1974.
- [10] P. Grossman, M. Spoerle, and F. H. Wilhelm, "Reliability of respiratory tidal volume estimation by means of ambulatory inductive plethysmography," *Biomed. Sci. Instrum.*, vol. 42, pp. 193–198, 2006.
- [11] A. Johansson and P. P. Å. Öberg, "Estimation of respiratory volumes from the photoplethysmographic signal. Part I: experimental results," *Med. Biol. Eng. Comput.*, vol. 37, no. 1, pp. 42–47, Jan. 1999.
- [12] Y. S. Lee, P. N. Pathirana, C. L. Steinfors, and T. Caelli, "Monitoring and Analysis of Respiratory Patterns Using Microwave Doppler Radar," *IEEE J. Transl. Eng. Health Sci.*, vol. 2, pp. 1–12, 2014.
- [13] G. Li, N. C. Arora, H. Xie, H. Ning, W. Lu, D. Low, D. Citrin, A. Kaushal, L. Zach, K. Camphausen, and R. W. Miller, "Quantitative prediction of respiratory tidal volume based on the external torso volume change: a potential volumetric surrogate," *Phys. Med. Biol.*, vol. 54, no. 7, pp. 1963–1978, Apr. 2009.
- [14] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, and others, "Standardisation of spirometry," *Eur. Respir. J.*, vol. 26, no. 2, pp. 319–338, 2005.
- [15] C.-L. Que, C. Kolmaga, L.-G. Durand, S. M. Kelly, and P. T. Macklem, "Phonospirrometry for noninvasive measurement of ventilation: methodology and preliminary results," *J. Appl. Physiol. Bethesda Md* 1985, vol. 93, no. 4, pp. 1515–1526, Oct. 2002.
- [16] O. Sayadi, E. H. Weiss, F. M. Merchant, D. Puppala, and A. A. Aroundas, "An Optimized Method for Estimating the Tidal Volume from Electrocardiographic Signals: Implications for Estimating Minute Ventilation," *Am. J. Physiol. - Heart Circ. Physiol.*, vol. 307, pp. H426–H436, 2014.
- [17] B. J. Semmes, M. J. Tobin, J. V. Snyder, and A. Grenvik, "Subjective and objective measurement of tidal volume in critically ill patients," *Chest*, vol. 87, no. 5, pp. 577–579, 1985.
- [18] R. Gilbert, J. H. Auchincloss, J. Brodsky, and W. Boden, "Changes in tidal volume, frequency, and ventilation induced by their measurement," *J. Appl. Physiol.*, vol. 33, no. 2, pp. 252–254, Aug. 1972.
- [19] B. A. Reyes, N. Reljin, and K. H. Chon, "Tracheal Sounds Acquisition Using Smartphones," *Sensors*, vol. 14, no. 8, pp. 13830–13850, Jul. 2014.
- [20] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote Measurements of Heart and Respiration Rates for Telemedicine," *PLoS ONE*, vol. 8, no. 10, p. e71384, Oct. 2013.
- [21] L. Sherwood, *Fundamentals of Human Physiology*, 4th ed. Boston, MA, USA: Cengage Learning, 2011.
- [22] M. Bartula, T. Tigges, and J. Muehlsteff, "Camera-based system for contactless monitoring of respiration," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2672–2675.
- [23] S. J. Cala, C. M. Kenyon, G. Ferrigno, P. Carnevali, A. Aliverti, A. Pedotti, P. T. Macklem, and D. F. Rochester, "Chest wall and lung volume estimation by optical reflectance motion analysis," *J. Appl. Physiol.*, vol. 81, no. 6, pp. 2680–2689, Dec. 1996.
- [24] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [25] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao, "Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 11, pp. 2760–2767, Nov. 2014.
- [26] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, p. 807, 2014.
- [27] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian Video Magnification for Revealing Subtle Changes in the World," *ACM Trans Graph*, vol. 31, no. 4, pp. 65:1–65:8, Jul. 2012.
- [28] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing," *J. Appl. Physiol.*, vol. 22, no. 3, pp. 407–422, Mar. 1967.
- [29] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.
- [30] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.
- [31] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 6, pp. 1461–1470, Dec. 1985.
- [32] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, and W. Jones, "Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison," *Signal Process.*, vol. 43, no. 2, pp. 149–168, May 1995.
- [33] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, pp. 077010–077010, 2011.
- [34] T. M. Baird and M. R. Neuman, "Effect of infant position on breath amplitude measured by transthoracic impedance and strain gauges," *Pediatr. Pulmonol.*, vol. 10, no. 1, pp. 52–56, 1991.
- [35] M. J. Tobin, S. M. Guenther, W. Perez, and M. J. Mador, "Accuracy of the respiratory inductive plethysmograph during loaded breathing," *J Appl Physiol*, vol. 62, no. 2, pp. 497–505, 1987.
- [36] V. P. Vellody, M. Nassery, W. S. Druz, and J. T. Sharp, "Effects of body position change on thoracoabdominal motion," *J. Appl. Physiol.*, vol. 45, no. 4, pp. 581–589, Oct. 1978.
- [37] W. S. Druz and J. T. Sharp, "Activity of respiratory muscles in upright and recumbent humans," *J. Appl. Physiol.*, vol. 51, no. 6, pp. 1552–1561, Dec. 1981.