On-Device Filter Design for Self-Identifying Inaccurate Heart Rate Readings on Wrist-Worn PPG Sensors

JUNGMO AHN1, (Member, IEEE), HO-KYEONG RA2, HEE JUNG YOON3, SANG HYUK SON4, (Life Fellow, IEEE), AND JEONGGIL KO5, (Senior Member, IEEE)
1Department of Artificial Intelligence, Ajou University, Suwon 16499, South Korea
2IBM Korea, Seoul 07326, South Korea
3School of Technology and Computing, City University of Seattle, Seattle, WA 98121, USA
4Department of Information and Computing Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, South Korea
5School of Integrated Technology, Yonsei University, Incheon 21983, South Korea
Corresponding author: Jeonggil Ko (jeonggil.ko@yonsei.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Korean Government [Ministry of Science and ICT (MSIP)] under Grant 2015R1A5A1037668, and in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program, supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2020-2020-0-01461.

ABSTRACT The ubiquitous deployment of smart wearable devices brings promises for an effective implementation of various healthcare applications in our everyday living environments. However, given that these applications ask for accurate and reliable sensing results of vital signs, there is a need to understand the accuracy of commercial-off-the-shelf wearable devices’ healthcare sensing components (e.g., heart rate sensors). This work presents a thorough investigation on the accuracy of heart rate sensors equipped on three different widely used smartwatch platforms. We show that heart rate readings can easily diverge from the ground truth when users are actively moving. Moreover, we show that the accelerometer is not an effective secondary sensing modality of predicting the accuracy of such smartwatch-embedded sensors. Instead, we show that the photoplethysmography (PPG) sensor’s light intensity readings are an plausible indicator for determining the accuracy of optical sensor-based heart rate readings. Based on such observations, this work presents a light-weight Viterbi-algorithm-based Hidden Markov Model to design a filter that identifies reliable heart rate measurements using only the limited computational resources available on smartwatches. Our evaluations with data collected from four participants show that the accuracy of our proposed scheme can be as high as 98%. By enabling the smartwatch to self-filter misleading measurements from being healthcare application inputs, we see this work as an essential module for catalyzing novel ubiquitous healthcare applications.

INDEX TERMS Heart rate monitoring, PPG sensor, reliable healthcare sensing, smartwatch, wearable devices.

I. INTRODUCTION

Smartwatches are now a ubiquitously deployed mobile device and can be considered a representative form of wearable platforms. Users wear smartwatches for various reasons. Some to quickly receive the smartphone’s notification alarms, some for fitness tracking and healthcare, and some to enjoy these new feature while simply keeping track of time. This work focuses on the fact that smartwatches, while not yet, hold the potential to act as a core components for at home and remote healthcare applications. Specifically, given that most smartwatches offer heart rate sensor readings they can be used to continuously monitor critical events that may occur to patients with various chronic cardiac disorders or even accurately track and quantify the activity levels of patients with maladies such as diabetes [2].

Unfortunately, unlike the hype in applying smartwatches to clinical and healthcare protocols, even the smartwatch vendors admit that the heart rate readings from smartwatches are only accurate under specific conditions [3]. This work
targets to design a subsystem that can be embedded within a smartwatch to track the validity of the heart rate measurements so that application designers can use the information to exploit measurements with enough confidence on its accuracy. Such a validation process is especially important for clinical and many healthcare applications given that the measurement accuracy leads to accurate health status analysis and decision/suggestion making [4].

To design a simple sensing system with small form-factor, heart rate measurements on smartwatches are usually done using a photoplethysmography (PPG) sensor. Specifically, PPG sensors on smartwatches use LED diodes to emit light to the skin, and an optical sensor measures the amount of light absorption at the veins to identify the precise times in which the heart beat generates blood flow. Nevertheless, despite its simplicity, accurate heart rate readings are challenging since the sensors can be sensitive to human movements and users wear the smartwatch in different ways. Some previous approaches target to remove the effect of such motion artifacts on the sensor readings using the smartwatch-embedded motion sensors (e.g., accelerometer/gyroscope) and a measurement calibration model, but with random user movements, such static approaches may not be effective enough. Nevertheless, the interest in applying smartwatches for healthcare applications request for some scheme to (at the very least) suppress inaccurate readings from being used as input to the application logic.

This work starts with a preliminary study phase in which we test the heart rate measurements collected from smartwatches of various manufacturers (c.f., Figure 1) with and FDA-approved Zephyr Bioharness chest strap as the ground truth measurements. Results from our preliminary study suggest that when the user stays still the accuracy of the heart rate measurements are very high with an error of only 1.23 bpm, but as the user actively moves the smartwatch heart rate measurements start to deviate from the ground truth. Based on such observations we design a light-weight system that can be embedded within smartwatches themselves. Specifically, we exploit the variations of the reflected LED signals from the skin captured at the photodiode, and use a Viterbi algorithm connected to a Hidden Markov Model to make predictions on the accuracy of the heart rate measurements. Overall our evaluations suggest that the proposed system correctly classifies the heart rate sensor readings with an accuracy of ~98% with ~8% false-negative ratios when the accuracy threshold is set to 5 bpm.

Specifically, the contributions made in this work can be summarized in three-fold.

- First, we present results on commercial-off-the-shelf (COTS) smartwatches’ capability in measuring heart rate under various wearing conditions to quantify the accuracy of sensor readings. Specifically, our experimental results show that the PPG sensor’s input light reading variance plays an important role in delivering accurate heart rate readings on mobile PPG sensors.
- Second, we propose a scheme to classify between accurate and inaccurate heart rate measurements using a Viterbi algorithm-based Hidden Markov Model. For designing such a scheme, we gather experiences from physicians to define tolerable heart rate measurement errors. The scheme presented here is designed to be light-weight so that they can be easily implemented on smartwatches themselves.
- Lastly, we implement the proposed algorithm on a COTS smartwatch, and present experimental results based on real-world data collected in both indoor and outdoor environments.

II. BACKGROUND AND PROBLEM STATEMENT

A. BACKGROUND

1) HEART RATE MONITORING

There are two different methods that are typically used for heart rate measurements: (1) the electrocardiogram (ECG) signals via electric signals or (2) photoplethysmography (PPG) samples made via optical measurements. ECG is the most commonly used method for clinical purposes in hospitals due to its highly accurate and reliable manner. ECG records and amplifies an electric signal generated from heart beats. For ECG-based heart rate measurements, lead cables and electrodes are attached to the subject’s chest near his or her heart. Given that ECG is captured using leads that are connected closely to the body, depending on the connectivity of the leads, it can be less-effectuated from motion artifacts. However, the form factor of ECG sensors are inappropriate for widely used mobile platforms. Therefore, on smartphones or smartwatches, PPG sensing is typically used. PPG sensors exploit LEDs that emits light to the skin and a photodiode that captures the reflected light from the skin. This process can be formulated by following equation [5], [6], the Beer-Lambart law which defines the attenuation of light to the properties of the penetrated material.

\[ I_o(t) = I_i \cdot \exp \left( -\Sigma_{j=1}^{n} \epsilon_{\lambda,j} \cdot c_j(t) \cdot d_j(t) \right). \]  

(1)

where \( I_i \) is the input light intensity of PPG sensor, \( \lambda \) is wavelength of the light, \( \epsilon_{\lambda,j} \) is the reflection/absorption coefficient
of the material (e.g., in our case, tissue of users’ skins), \( c \) is concentrations of the material, \( d \) is the light path length, and \( I_o \) is the intensity of reflected light which the result of absorption and reflection traveling the blood flow. As we aforementioned, the photodiode captures the \( I_d \) signal to measure heart rate. The wavelength of the emitted LED on a PPG is between 500 nm to 600 nm, which represents the green-yellow region of the visible light spectrum [7]. If the wavelength is longer (e.g. 630nm; red light region), the reflected light intensity becomes weaker since longer wavelengths penetrate the tissue of skin deeper than shorter wavelengths. Thus, the use of green-yellow wavelengths allows for less impact from motion artifact compared to other light wavelengths [8].

The PPG sensor can also be designed in two ways, where in the first the reflectance of light is utilized (as discussed until now) and in the second we use a “transmission mode”. The difference between the two is mainly in the location of the photodiode. When exploiting the transmission mode, the LED and a photodiode are located at opposite sides of the body (e.g., earlobe, finger). Thus, the photodiode reads LED light passing through the body parts. This approach relatively shows better results [7]. However, this approach can only be applied to specific body parts; thus, cannot be used in general platforms. On the other hand, when using the light reflectance, the photodiode can be located close to the LED and the form factor can be minimized. Unfortunately, this approach is heavily affected by motion artifacts. Nevertheless, since reflectance-based PPG sensors are most widely used we focus on these devices for the remainder of this work.

2) PPG SENSORS WITH MOTION ARTIFACTS

As aforementioned, reflectance-based PPG sensors are used on many commercialized mobile wearable devices due to its compact design. However, even the simplest hand gestures can cause inaccurate sensor readings due to motion artifacts. Specifically, such motion artifacts will impact expressions \( c \) and \( d \) in Equation 1. When the location of LED and photodiode is moved, the concentration of material \( c \) changes and when the space between skin and both LED and photodiode impacts \( d \), the light path length [6]. Thus, when the user is mobile, the photodiode is likely to read abnormally reflected light \( I_d \) from the user’s skin. Consequently, motion artifacts debase the quality of the PPG signal [4].

While it is difficult to completely remove motion artifacts or reconstruct the original PPG signals, many studies have tried to minimize the impact of motion artifacts from wearable PPG sensors [9]–[13]. These studies exploit physical information extracted from an accelerometer to perform noise reduction, signal reconstruction and heart rate estimation. We will discuss these efforts in greater detail using Section III.

B. PROBLEMS

- Ubiquitous usage: Mobility is a major advantage that PPG sensor-equipped wearable devices possess compared to using a static ECG monitor. It is known that 45% of users bought a smartwatch to use as activity trackers [14] as a way to exploit such advantages. Furthermore, a wearable device can offer heart rate measurements in a miniature form-factor, and mobile device manufacturers even produce earbuds which can monitor the user’s heart rate using the PPG sensor. However, paradoxically, PPG sensors cannot assure high reading reliability due to the user’s mobility. We will further formulate this problem through our preliminary study (Section IV).

- Inapplicable measurement reconstruction: There have been a number of previous studies that propose PPG signal and heart rate estimation reconstruction. These works showed the feasibility of heart rate reconstruction and achieved reasonable performance. However, we note that clinical sensor signal noise reduction and reconstruction can cause negative implications to clinical diagnose; thus, cannot be directly used for clinical use. Rather, simply identifying what samples are right an what are wrong can give more practical feedback to the clinical personnel [15].

- Healthcare application design: Wearable devices have hardly been considered as clinical-grade devices due to the unreliable performance of PPG sensors under motion artifacts. Nevertheless, wearable devices with PPG sensors still hold the potential to be used as clinical devices because of their mobility. For example, in hospitals, fingertip PPG sensors are attached to patients’ finger to collect pulse rate and \( \text{SpO}_2 \) blood oxygen saturation from a bedside monitoring system. Wearable devices are also attached to users’ skin so they also can work similarly like fingertips connected to bedside monitoring system. For example, Apple designed an arrhythmia detection application for the Apple watch using a PPG and ECG sensor, respectively [16]. Even though wearable devices have such potentials, developing clinical applications (e.g., monitoring heart arrhythmia) is not easy if the wearable device cannot assure the reliability of heart rate readings. Thus, there is a need to at the very least identify which measurements are accurate (or reliable) and which are not.

III. RELATED WORK

Previous studies have examined how to handle the impact of motion artifacts on the PPG sensor installed in wrist-worn devices. Most of these previous work employed various signal processing algorithms to reconstruct abnormal raw PPG sensor signals and also try to estimate the heart rate from the reconstructed PPG signal. Zhang et al. [11], [17] proposed pre- and post-processing algorithms to achieve an average estimated error of 2.34 bpm [11] using signal decomposition for eliminating noise from motion artifacts, sparse signal reconstruction, and spectrum peak tracking. The authors achieve 1.28 bpm estimated error using joint sparse spectrum reconstruction using the multiple measurement vector model.
and spectral subtraction. Biswas et al. [18] proposed a deep learning-based approach. Using two convolutional neural networks (CNN) layers, two long-short term memory (LSTM) layers, and a dense layer, their approach achieves 1.47 bpm average estimation error for data collected from 20 subjects. Roy et al. [19] proposed artificial neural network which improves heart rate measurement under motion artifacts and shows high correlation 0.99 against a ECG-based heart rate measuring device.

However, the aforementioned previous works require high computation power and cause latency; thus, making it difficult to address real-time constraint for continuous PPG inputs. Especially, the work by Zhang et al. [11] spends 0.942 seconds for reconstructing one time window slot (8 seconds) [20] using a PC with Intel Core 2 Duo P7550 @2.26GHz and 4 GB RAM.

Sun et al. propose a low latency heart rate estimation scheme for samples collected under high motion artifacts using asymmetric least squares spectrum subtraction and Bayesian decision theory [20]. This approach shows 2.13 bpm average estimation error and spends 0.0162 seconds for estimating 8 seconds of data (using the same hardware in [20]). Chung et al. also proposed a fast heart rate estimation algorithm based on a finite state machine with a crest factor and heart rate change [21]. They show that the proposed algorithm only requires 1.1 ms for estimating 8 seconds of data (using Intel Core i7-3770 CPU@3.40 GHz) while providing less than 1 bpm average estimation error from 23 subject datasets.

In recent studies, multi-wavelength a PPG sensor design is suggested in order to mitigate and reduce the impact of motion artifacts by combining multiple signals. Ishikawa et al. propose a new PPG sensor design embedded on a wristband form factor device that measures heart rate robustly under motion artifacts using two different wave-length LEDs; green LED for arm-motion noise reduction and red LED for finger-motion noise reduction [8]. By combining accelerometer and PPG signals, their proposed scheme can instantly reduce incorrect measurement. Zhang et al. suggest the motion artifact reduction algorithm using green PPG signal and IR PPG signal using wavelet transform and signal reconstruction. Lee et al. suggest combining 12 channel PPG signals generated by four PPG sensors each containing red, green and infrared LEDs [22].

We note that most of the previous works refer to the accelerometer signal for handling motion artifacts or integrate multiple PPG sensor signals. However, in this paper, we discuss why the accelerometer-based approach might not be sufficient to measure the reliability of heart rate measurement directly on a smartwatch. Furthermore, a novel PPG sensor design is an attractive approach but such methods cannot be applied to the many already-deployed PPG-based heart rate sensors and even for new devices, may increase the production cost. Furthermore, we point out again that using the heart rate estimation from a broken PPG signal may cause negative implications for healthcare applications. We also point out that such reconstruction algorithms can cause heart rate measurement delay when they are operated on wearable devices even if the algorithm can operate very fast on PC class devices.

IV. PRELIMINARY STUDY
In this preliminary study, we observe the difference of heart rate sensor readings when the user is staying still or naturally walking. Furthermore, we investigate into the root cause of misleading heart rate measurements and what the requirements for designing a filter for these misleading measurements are.

A. DATA COLLECTION
For data collection, we use three smartwatches from different manufacturers; Samsung Gear S2; LG Watch Urbane; Apple Watch. These devices are based on three different operating systems, Tizen, Android Wear OS and Apple WatchOS, respectively. We implement applications for sensor data logging on each OS platform. This application is loaded to the smartwatch to record the heart rate measurements and 3-axis acceleration. The heart rate measurements are taken every second and the accelerometer sampled at a rate of 1 KHz. For capturing ground truth heart rate measurements, we employ Zephyr’s Bioharness device which is chest-band type heart rate logger using its two-lead ECG sensor, which is more robust to user movements than the PPG sensors used on the smartwatch devices. The Bioharness, approved by the FDA, allows accurate heart rate measurements and we implement an Android platform-based sensor data logging application to extract its data.

For measuring the heart rate from the sensors, we implemented an application for logging the heart rate from the smartwatches we tested with. Specifically, for the Apple Watch, the application was implemented using Apple HealthKit and the application is designed to sample heart rates based on the Apple watchOS configurations (given that Apple Watch does not allow modification to the sampling interval) and the measurement is sent to an iPhone application via Bluetooth connections.

Fortunately, for the LG Urbane application developed using Android Wear, we were able to set the sampling rate to 1 Hz, and this application performs similar behavior of sending its heart rate measurements (on a per-second basis) to its associated smartphone app using Bluetooth.

Finally, the Samsung Gear S2 smartwatch operates Tizen OS. Since the Gear S2 is equipped with an on-board WiFi chipset, the application is designed to transmit the heart rate measurement (sampling rate of 1 Hz) to our data analytics server directly. Furthermore, for future use, for the Gear S2, we also capture and report the light intensity measurements of the PPG sensor (not provided in Apple HealthKit of Android Wear) for each heart rate measurement as well. Unlike watchOS and Android Wear, in which the raw measurements of heart rate signals pass through software calibrations, Gear S2 allows access to the raw heart rate value.
measurements directly from the PPG sensor, and we report this raw value to the server.

**B. SMARTWATCH HEART RATE ERRORS**

With the software designed for each smartwatch platform, the initial target of our study was to understand how the motion and wearing patterns of smartwatches impact the heart rate measurement quality. For capturing ground truth heart rate data (for comparisons), we use the Zephyr BioHarness device, which is an FDA-approved chest-strap device for physiological signal measurements. Given its form-factor, the BioHarness devices are more robust towards user motions and thus, has been applied to many previous healthcare applications [23]–[25]. With the smartwatches and the BioHarness device, we recruit four volunteers (3 male, 1 female; avg. age 24) and ask them to make natural walking behaviors in two different environments: (i) hallway under florescent light and (ii) an open field under natural day-light. The participants were wearing the BioHarness device on their chest and different smartwatch devices were worn over multiple test runs. We note that while the walking speeds of each participant was not fixed and we intentionally did not guide the participants to move their arms in specific ways, the walking speeds were in the ranges of 1.2-1.8 m/s, and there were no cases in which the arm postures were fixed. We do so to capture natural walking conditions from the participants, rather than generating an artificially configured dataset. Note that the participants wore each watch twice: once comfortably loose on their wrist (natural wearing pattern) and in the second turn, very tightly.

![Figure 2](image_url)

FIGURE 2. Heart rate readings for the BioHarness (purple; ground truth) and three COTS smartwatches (green) while continuously moving. The plots also present the smartwatches’ internal accelerometer readings (in light blue, orange, yellow for each axis) and the PPG sensor light intensity plots captured from the Samsung Gear S2 watch (blue).

The plots in Fig. 2 suggests some interesting points. First, we notice that when the user tightly wears the watch, the smartwatch readings are highly correlated with the ground-truth. Quantitatively, the error (or the average difference with the ground-truth) is 3.4 bpm. This implies that with proper adjustments, the smartwatch’s heart rate readings can be very accurate. On the other hand, second, when the watch is worn loosely, we see a much higher error (avg. 18.4 bpm difference from the ground truth; Apple Watch–8.73, Urbane–17.78, and Gear S2–28.69). This suggests that human motion itself is only a critical factor that reduces the accuracy of heart rate measurements when the wearing conditions of the watch is not tightly attached to the human skin. Our discussions with physicians at hospitals and the ANSI/AAMI EC13 standards indicate that heart rate reading errors of ±5 bpm or ±10% bpm is tolerable for clinical use [26]. Third, we noticed that the accelerometer patterns were not sufficient enough to distinguish between the cases in which the watch was worn tightly or loosely, or in other words, it could not be an effective indicator on the potential accuracy of the heart rate readings. Quantitatively, by comparing the cases in which the watch was worn tightly and loosely (with lag adjustments), the accelerometer readings showed a mean cross-correlation coefficient of 0.77. Such finding suggests that using the accelerometer (alone) cannot be a good indicator of the heart rate measurement accuracy. Lastly, we can notice from plots in Fig. 2 (c) and (f), that the light intensity readings show noticeably different patterns for the tightly worn and loosely worn cases. This can be an indicator that the light intensity readings can be an effective feature for determining between the two cases and the smartwatch’s heart rate reading accuracy.

Based on this preliminary findings, we present results from additional experiments designed to validate the impact of different factors that affect the smartwatch PPG sensor’s light intensity measurement, which in turn, impacts the accuracy of heart rate readings. In Fig. 4 we present four different types of 3D printed rings that are designed to fit between the
smartwatch and the wrist. Specifically, these four rings were designed to validate the impact of distance changes from the PPG sensor with the human skin, thus, the absolute value differences of light intensity (i.e., first three rings with different heights), and also to validate the impact of external light with the same distance (i.e., 5mm ring with and without holes) by introduce externally impacted variance. For the absolute light intensity experiment, referring back to Equation 1, given the added distance of 3, 5, and 7 mm between the sensor and the skin, the distance of reflective light path, $d$ will be impacted through these changes.

In Fig. 3 we compare the heart rate readings errors observed from the three watches of our interest. We can see from Fig. 3 (a) that when standing still, the errors for all cases were kept low (less than 10 bpm for all test cases). Nevertheless, we can see a higher error (for the LG Urbane) when the distance from the skin to the PPG sensor increases to 7mm or when holes in the intermediate ring introduces external light variations. Still, these results suggest that the smartwatch can produce reliable sensor readings when the participant in still conditions caused for human arm motion is a more significant factor affecting the heart rate measurement quality. Based on this observation, in the following section we present a filter design which exploits such findings to identify the reliability of individual heart rate measurements on the smartwatch.

V. APPLICATION DESIGN

In this section, we discuss how to design a filter for determining the reliability of heart rate measurements on a smartwatch. We present the design goal of the filter and present the design of an HMM-based machine learning model. We conduct preliminary evaluations of the filter and we show how the performance of the filter can be improved by applying additional smoothing techniques. Finally, we show an application-level implementation of our filter on a smartwatch platform.

A. DESIGN GOALS

The following are the design goals we focused on when designing a filter for classifying accurate heart rate measurements on smartwatches.

- **Maintaining low false positive rates**: For the sake of potentially applying our filter to clinical applications, the filter should maintain a low false positive rate. Achieving a low false positive rate means that the filter can well classify inaccurate measurements as inaccurate. False positive rates can differ depending on the threshold which determines whether the measurement is accurate or not. Thus, if the threshold is lenient (e.g., 3 bpm), the false positive rate naturally decrease because the model conservatively classifies the sample data as accurate, but, in turn, the true negative rate can increase. This is a trade-off that should be taken into account when designing specific parameters for the filter.

- **On-device operation**: The goal of this work is operating the proposed filter on a smartwatch. To achieve this goal, we should consider three issues: energy efficiency, real-time operations, and low computation overhead. If the filter requires complicated algorithms as in previous heart rate estimation work we have discussed in Section III, it is difficult to meet such resource constraints. For example, signal decomposition and reconstruction algorithms for eliminating motion artifact induced noise can require approximately 1 second to estimate heart rate from 8 point of PPG signal data [11]. Moreover, if a series of PPG signal is transmitted directly to the server for detailed estimation, the latency will inevitably increase due to the transmission overhead. Thus, the filter’s decision-making process should consist of a light-weight machine learning algorithm.

- **Easy integration with healthcare apps**: Our proposed filter should be easily used by healthcare application developers. We believe that an easy-to-use API should

![Figure 3](image-url) Average heart rate error with different ring thicknesses for different smartwatch devices.

![Figure 4](image-url) 3D printed rings with different thickness: 3mm, 5mm, 7mm and 5mm ring with holes.
be offered so that linking our filter to an application is made simple. For now, the only feedback that the heart rate sensors offer to the application is whether the sensor is currently connected to the user’s skin or not. The proposed filter should offer additional information on the accuracy of the heart rate measurements to the application.

B. FILTER DESIGN

Results from our preliminary results suggest that we can potentially design a filter that classifies a smartwatch’s heart rate reading accuracy by understanding and analyzing the light intensity variance patterns of the PPG sensor. A simple approach to do so would be to configure a threshold on the variance measurements, but based on our (failed) efforts, the changes between consecutive readings were overly rigorous and setting a static threshold was difficult to achieve.

The filter design is based on the Viterbi algorithm to seek the most possible sets of hidden states. In our application, we consider the sensing data quality as the hidden states in the Viterbi algorithm. Specifically, we identify the Viterbi path that outputs a sequence of observed events. We select to use such an approach given that the PPG sensor’s light intensity measurements are captured as the observations for the algorithm.

We start describing the design of our filter by presenting its HMM structure. A typical HMM structure consists of states, observations, and probabilities. The state is hidden to the observer and only can be estimated by using computed probabilities. Fig. 5 depicts our HMM filter design.

\[ State \ St = \{ \text{Good}, \text{Bad} \} \]

\[ Observation \ O = O_1, O_2, \ldots, O_n \]

In our filter design, we define two different states: reliable sample (Good) and non-reliable (Bad) sample. The initial state is determined based on the start probability (solid lines in Fig. 5). The state transition probability (rounded dotted lines in Fig. 5) is a probability to which a state \( St_i \) in time \( t_n \) transits to a state \( St_j \) in time \( t_{n+1} \). The emission probability (downwards dotted lines in Fig. 5) is a probability that observation \( O_n \) will be observed in each state \( St \). Since the initial HMM states are critical in the decision making of accurate estimations, the HMM is trained using the Baum-Welch expectation maximization algorithm via a sequence of samples with different light intensity measurements. The output is selected by the state based on observations. Observations are used to compute the corresponding path at each point of time.

An observation is generated by flooring the change of light intensity over step size. For both of training and testing, multiple observations are fed into initial and trained model to identify the current state. Note that the data collected from the Samsung Gear S2 watch presented in Fig. 2 shows that as the heart rate readings diverge from the ground truth readings, we see more variations in the light intensity measurements. Based on such an observations, we define the set of measurement differences among two consecutive light intensity measurements for a time window \([t : t + w]\) as \( \Delta L_{t:t+w} \). We also denote the absolute maximum value in this set as \( \max(|\Delta L_{t:t+w}|) \). In our algorithm, we start by dividing \( \max(|\Delta L_{t:t+w}|) \) with \( N_O \), the number of possible observations to extract \( S_{\text{step}} \), the step size.

\[
S_{\text{step}} = \frac{\max(|\Delta L_{t:t+w}|)}{N_O} \tag{2}
\]

We configure \( N_O \) by analyzing the light intensity variance traces. Based on the data we collected, we noticed that a difference of more than 6000 units for the light intensity caused significant variations in heart rate measurements. The min-max ranges of the PPG sensor light intensity suggests that for our data set \( N_O \) should be set to 11. We plan to design algorithms to adaptively configure such parameters with respect to the input data as part of our future work since the number of possible observations can impact the granularity and responsiveness of the system.

Notice that we maintain two states in the HMM, one for declaring an accurate reading and the other for identifying inaccurate measurements. At each time instance in which we compute the heart rate measurement accuracy, we compute an observation \( O_n \) (e.g., HMM input) as the following.

\[
O_n = \left\lfloor \frac{\max(|\Delta L_{t:t+w}|)}{S_{\text{step}}} \right\rfloor \tag{3}
\]

Fig. 6 plots a time-series of a sample collected from the Samsung Gear S2 and the BioHarness (ground-truth device) with comfortable (loose) tightness. In the experiment, the user started with 3-5 minutes of “not moving” followed by \( \sim 10 \) minutes of actively moving. The study participants and movement conditions were kept identical to the experiments in Section IV-B. As expected, as the user starts to move, the accuracy of the smartwatch readings start to diverge from the Bio-Harness. The black regions in this figure represent the time instances when the ground-truth and the smartwatch

\[ FIGURE 5. \text{HMM model with two states and 11 observations. Each line depicts probabilities, where the solid, dashed, and dotted lines indicate the start probability, transition probability, and emission probability, respectively.} \]
readings differ by more than 5 bpm and the blue region presents our filter’s classification result classified as ‘inaccurate sample’ at a given time. Thus, ideally, the black region and the blue region should show identical patterns if the proposed filter functions perfectly. We can notice that for the majority of the cases, our proposed filter properly makes predictions on the accuracy of the smartwatch’s heart rate sensor. Nevertheless, a small number of false predictions still persist. Quantitatively, in this sample experiment, the accuracy of detecting samples with differences of more than 5 bpm by our proposed scheme was ~90%. A deeper look into the light intensity readings from the PPG sensor suggest that their measurements cannot be perfectly reliable. Since the fundamentals of our filter relies on the light intensity value variations, such noise can lead to mis-understanding of the data.

We address such limitations by focusing on the fact that physiologically, heart rate measurements show slow changing patterns given that the measurements themselves are an average value over a minute. Thus, we apply a moving average on the input samples to smoothen the impact of outliers and use this value as the observation input. As Fig. 7 shows, the modified filter design effectively suppresses wrong classification results when the measurements from the smartwatch diverge from the ground truth. Quantitatively, the accuracy of predicting samples that are inaccurate (with more than 5 bpm difference from the ground truth) is increased to ~92% for this sample trace when adding moving averages to our scheme.

C. APPLICATION IMPLEMENTATION

The implementation of our heart rate measurement accuracy filtering module is to output two different values for the applications to exploit. First, the heart rate measurements from the smartwatch, and second, the reliability level of the current measurement. In Fig. 8 we present the functional diagram and a simplified code implementation of our filter. Here, the main function of our filter `reliable_heart_rate_reading` creates a queue Q to store a time window W number of light intensity measurements. Once W light intensity measurements are captured in Q, our model outputs an observation using the observation function. This function is essentially an implementation of Equation (3) for computing an observation O as an input of the HMM model. This observation is then used as input to the HMM filter function, which operates the HMM model and returns the reliability of the current observation O. By outputting both the heart rate value and the measurement reliability, applications that connect to our filter can selectively use heart rate measurements based on the expected accuracy of the samples.

Note that `reliable_heart_rate_reading` takes in a threshold value as its parameter. This threshold represents the tolerance value in which the measured heart rate value can differ from the ground truth. The threshold parameter represents the tolerance value of the difference between the ground truth and Samsung Gear S2 measurements when training the model. Thus, we train for models with different...
threshold values and apply the proper filter when in operation. Specifically, a low threshold suggests that the filter strictly determines that the heart rate measurement is reliable only when the measurements are very accurate. However, such a low threshold filter may return fewer heart rate measurements marked as “reliable” given that many samples can be classified as “unreliable”. If the application developer determines that the application requires highly reliable heart rate measurements, the a low threshold can be configured. On the other hand, if the application does not require highly reliable heart rate measurements, a high threshold can be set and samples that are not too different from the ground truth will be offered to the application.

VI. EVALUATION
We now evaluate the proposed filter design in two aspects: (1) performance of filtering unreliable heart rate measurements, and (2) latency and battery consumption with respect to the filter operations. We use the Samsung Gear S2 smartwatch platform for our experiments which is the only platform that supports PPG sensor’s light intensity recordings among the devices of our interest. We point out that the HMM model training is done on a PC class device and implanted the HMM model to the smartwatch.

A. FILTER PERFORMANCE
Figure 9 shows the overall classification results of our proposed filter design for varying target accuracy thresholds (e.g., an accurate estimation determined if the observed sample has less than 3, 5, 7, 10 bpm difference from the ground-truth). Here, we use a 100 minute data set (25 minutes of data for each of the four volunteers) which included a mixture of walking and still data. Results in Figure 9 indicate that the rate of properly classifying “accurate” samples as “accurate” (i.e., true-positive) is >90% for all cases in which the accuracy threshold exceeds 5 bpm. The false-positive cases, which is the ratio of incorrectly classifying accurate samples as inaccurate, can be kept low in such cases. As a system to be used in healthcare applications, it is more important to properly prevent inaccurate samples from being used as system-level inputs. Thus, it is important to minimize the false positive rates. Our system, with proper accuracy configurations, show a low false positive rate (8% when with a 5 bpm threshold and 4% with 3 bpm threshold), making it suitable for real application use.

Overall, our system design and evaluation results, suggest that we can accurately classify the accuracy of a smartwatch’s heart rate measurements with the complexity that can operate on the smartwatch internally.

B. LATENCY AND ENERGY CONSUMPTION
While being an effective back-end system for improving the application quality, we should make sure that the utilizing our filter does not negatively affect the system performance. Given that the algorithm operates on severely resource limited platforms, observing the system-level performance impact is even more important. For this, in this section, we present the latency and battery consumption of using our filter on a smartwatch. Specifically, we measure the elapsed time and power usage overhead for operating the filter on the Samsung Gear S2 smartwatch, which is equipped with a Exynos 2 Dual 3250 CPU (7.2 GFLOPS). The filter operation is consists of reading heart rate measurements using the system API, computing the observation from the HMM model, and operating the filter to retrieve the current state of PPG sensor’s output value.

Using 1,500 samples output from the filter, we were able to notice that the average latency for computing the output of a sample is 0.2 msec (±0.08 msec). Given that the smartwatch offers heart rate measurements of 1 Hz [27], we see this latency acceptable for practical applications [28].

For measuring the battery usage we use the Dynamic Analyzer offered by the TIZEN developers group [29] and measure the average current draw in mA units for three minutes. For references, we also test for cases in which heart rate monitoring does not take place, and also for a case where the PPG sensor takes heart rate measurements, but with no filter operations. Fig. 10 plots the results from these two cases along with the case when our filter is used. Notice there that the baseline current draw (when not using the PPG sensor with the app turned on) is 0.18 mA (±0.03). Enabling the PPG sensor itself adds a significant amount of current draw overhead to consume 2.74 mA (±0.58) on average. Finally, applying our filter added an additional 1 mA by resulting in 3.71 mA (±0.43). Note that these numbers are only for the application that we designed implemented for testing and

\[ \text{FIGURE 9. Overall classification results of our proposed filter on 100 minute data set with different accuracy thresholds.} \]

\[ \text{FIGURE 10. Average CPU power estimation of HR monitoring application implemented on target device Samsung Gear S2.} \]
does not include the current draw caused from other system components, which as the OS, display, networking, etc. We note that the system’s baseline current draw is \( \sim 42.8 \, mA \) when running an idle app. Therefore, the 1 mA increase due to the use of our filter translates to only 2.3% of the entire system’s current draw.

**VII. CONCLUSION**

Wearable devices have become more and more prevalent and can provide the potential to enable various healthcare applications by tracking and recording a user’s daily health status-related information. In spite of the such promising potential of enabling important applications, we noticed that the performance of heart rate measurements on wearable platforms may not be reliable due to unavoidable motion artifacts. In this paper, we focus on the fact that such motion artifacts heavily (and negatively) impact the heart rate measurements that are collected from PPG sensing units. Based on preliminary studies, we show that there are significant variations on the PPG sensor’s light intensity readings when mixed with external motion artifacts; leading to inaccurate heart rate measurements. To resolve this issue, we propose a Hidden Markov Model-based filter design to determine the reliability of each heart rate reading. Our evaluations with the proposed filter shows a classification accuracy of 98%. Especially we noticed that the false positive rate which can potentially cause negative implications on healthcare or clinical decisions is only \( \sim 8\% \) when configuring 5 bpm as the threshold for determining accurate measurements. Finally, using the proposed filter, we implement a heart rate monitoring application on a COTS wearable device. Our application computes the reliability of the current PPG sensor with minimal latency and with 2% overhead in current draw.

**ACKNOWLEDGMENT**

An earlier version of this article was presented by Ra et al., “I Am a Smart Watch, Smart Enough to Know the Accuracy of My Own Heart Rate Sensor” at the ACM International Workshop on Mobile Computing Systems and Applications in 2017.

**REFERENCES**

[1] H.-K. Ra, J. Ahn, H. J. Yoon, D. Yoon, S. H. Son, and J. Ko, “I am a ‘smart’ watch, smart enough to know the accuracy of my own heart rate,” in *Proc. 18th Int. Workshop Mobile Comput. Syst. Appl.*, 2017, pp. 49–54.

[2] P. Bonato, “Wearable sensors/systems and their impact on biomedical engineering,” *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 18–20, May 2003.

[3] P. Larkin, *Mio Boss: Fitbit and Apple are Getting Heart Rate Monitoring Wrong*, Accessed: Oct. 21, 2020. [Online]. Available: http://www.wareable.com/meet-the-boss/mio-boss-fitbit-apple-heart-rate-monitoring-957

[4] J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007.

[5] Y. Zhang, S. Song, R. Vullings, D. Biswas, N. Simões-Capela, N. van Helleputte, C. van Hoof, and W. Groenendaal, “Motion artifact reduction for wrist-worn photoplethysmograph sensors based on different wavelengths,” *Sensors*, vol. 19, no. 3, pp. 673, Feb. 2019.

[6] P. Renevey, R. Vetter, J. Krauss, P. Celka, and Y. Depeursinge, “Wrist-located pulse detection using IR signals, activity and nonlinear artifact cancellation,” in *Proc. Conf. Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, Oct. 2001, pp. 3030–3033.

[7] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, “Photoplethysmographic sensors—Past and present,” *Electronics*, vol. 3, no. 2, pp. 282–302, Apr. 2014.

[8] T. Ishikawa, Y. Hyodo, K. Miyashita, K. Yoshifujii, Y. Komoriya, and Y. Imai, “Photoplethysmograph PPG sensor for instant heart rate in daily activity,” in *Proc. 10th Int. Joint Conf. Biomed. Eng. Syst Technol.–BIOSIGNALS (BIOSTEC)*, vol. 4. SciTePress, 2017, pp. 126–133, doi: 10.5220/000610901250133.

[9] 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. 1885, 2010.

[10] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and A. K. 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.

[11] Z. Zhang, Z. Pi, and B. Liu, “TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.

[12] D. Dao, S. Salehzadeh, Y. Noh, J. Chong, C. Cho, D. McManus, C. Darling, Y. Mendelson, and K. Chon, “A robust motion artifact detection algorithm for accurate detection of heart rates from photoplethysmographic signals using time-frequency spectral features,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1242–1253, Sep. 2017.

[13] H. Lee, H. Chung, and J. Lee, “Motion artifact cancellation in wearable photoplethysmography using gyroscope,” *IEEE Sensors J.*, vol. 19, no. 3, pp. 1166–1175, Feb. 2019.

[14] K. Gustafson, “Smartwatch or Fitness Tracker? Why Age, Sex Matter,” Accessed: Oct. 11, 2020. [Online]. Available: https://www.cnbc.com/2015/01/07/smartwatch-or-fitness-tracker-why-age-sex-matter.html

[15] J. Park, W. Nam, J. Choi, T. Kim, D. Yoon, S. Lee, I. Paek, and J. Ko, “Glasses for the third eye: Improving the quality of clinical data analysis with motion sensor-based data filtering,” in *Proc. 15th ACM Conf. Embed. Netw. Sensor Syst.*, Nov. 2017, pp. 1–14.

[16] Apple. Using Apple Watch for Arrhythmia Detection. Accessed: Oct. 11, 2020. [Online]. Available: https://www.apple.com/healthcare/site/docs/Apple_Watch_Arrhythmia_Detection.pdf

[17] Z. Zhang, “Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1902–1910, Aug. 2015.

[18] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharya, C. Van Hoof, M. Konijnenburg, and N. Van Helleputte, “CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 2, pp. 282–291, Apr. 2019.

[19] M. S. Roy, R. Gupta, J. K. Chandra, K. D. Sharma, and A. Talukdar, “Improving photoplethysmographic measurements under motion artifacts using artificial neural network for personal healthcare,” *IEEE Trans. Instrum. Meas.*, vol. 67, no. 12, pp. 2820–2829, Dec. 2018.

[20] B. Sun and Z. Zhang, “Photoplethysmography-based heart rate monitoring using asymmetric least squares spectrum subtraction and Bayesian decision theory,” *IEEE Sens. J.*, vol. 15, no. 12, pp. 7161–7168, Dec. 2015.

[21] H. Chung, H. Lee, and J. Lee, “Finite state machine framework for instantaneous heart rate validation using wearable photoplethysmography during intensive exercise,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1595–1606, Jul. 2019.

[22] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, “Wireless sensor networks for healthcare,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, p. 1493, Mar. 2020.

[23] J. Juen, Q. Cheng, and B. Schatz, “Towards a natural walking monitor for pulmonary patients using simple smart phones,” in *Proc. 5th ACM Conf. Bioinf., Comput. Biol., Health Informat. (BCHB)*, 2014, pp. 53–62.

[24] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Tezis, and M. Welsh, “Wireless sensor networks for healthcare,” *Proc. IEEE*, vol. 98, no. 11, pp. 1947–1960, Nov. 2010.

[25] A. Kokonoz, A. Astoras, P. Semertzidis, E. Michail, D. Filos, I. Chouvarda, O. Grossenbacher, J.-M. Koller, R. Leopoldo, J.-A. Porche, M. Correvon, J. Luprano, C. Zamboulis, A. Sipila, and N. Maglaveras, “Development and clinical evaluation of a physiological data acquisition device for monitoring and exercise guidance of heart failure and chronic heart disease patients,” in *Proc. Comput. Cardiol.*, Sep. 2010, pp. 1099–1102.
JUNGMO AHN (Member, IEEE) received the B.S. degree in software and computer engineering from Ajou University, Suwon, South Korea, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Artificial Intelligence. His research interests include designing embedded learning systems, embedded deep learning systems for classifying compressed wireless sensor images and capsule endophathy images, and a low-cost IoT sensing platform for VOC material classification. He is a member of the Embedded Intelligent Systems (EIS) Laboratory, Yonsei University, leading by Dr. J. Ko. He is also a member of the Computer Systems Laboratory (CSL), leading by Dr. J. Ahn. He was a recipient of the Global Ph.D. Fellowship from the National Research Foundation of Korea.

HO-KYEONG RA received the B.S. degree in computer engineering and systems from the University of Washington, Tacoma, WA, USA, in 2011, and the M.S./Ph.D. degree in information and communication engineering from the Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, South Korea, in 2018. Since 2019, he has been a Senior Consultant with IBM Korea, Seoul, South Korea. He has collaborated on various smart home projects as well. His research interests include designing and developing smart medical and healthcare applications, automatic cell-counting, asthma monitoring and advice ecosystems, accurate heart rate monitoring, and cloud framework for smart healthcare systems. He was a recipient of the Excellent Researcher Award of the Graduate School at the DGIST in 2016.

HEE JUNG YOON received the B.S. degree in health informatics/health information management from the University of Washington, Seattle, WA, USA, in 2010, and the M.S. degree in information and communication engineering and the Ph.D. degree in information and communication engineering from the Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, South Korea, in 2013 and 2018, respectively. She has collaborated on various healthcare projects as well. In 2018, she was a Postdoctoral Researcher with the Cyber-Physical Systems Global Center, DGIST. Since 2019, she has been an Associate Faculty with the City University of Seattle, Seattle. Her research interest includes designing and developing surveillance applications for vulnerable people using a 3D depth camera fall detection for elderly individuals, aggressive behavior recognition for dementia patients, and child versus adult classification for children’s safety. She was a recipient of the DGIST Presidential Fellowship Award in 2016.

SANG HYUK SON (Life Fellow, IEEE) received the B.S. degree in electronics engineering from Seoul National University, the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), and the Ph.D. degree in computer science from the University of Maryland, College Park, MD, USA. He has been a Professor with the Department of Computer Science, University of Virginia, and the WCU Chair Professor with Sogang University. He has been a Visiting Professor with the KAIST, the City University of Hong Kong, Ecole Centrale de Lille in France, Linkoping University, and the University of Skövde, Sweden. He is currently a Chair Professor of the Daegu Gyeongbuk Institute of Science and Technology (DGIST), where he served as the Dean of the Graduate School and the President. He has authored or coauthored over 350 articles and edited/authored four books in his research areas. His research has been funded by the Korean Government, the National Research Foundation, the National Science Foundation, DARPA, the Office of Naval Research, the Department of Energy, the National Security Agency, and IBM Korea. His research interests include cyber physical systems, real-time and embedded systems, database and data services, and wireless sensor networks. He is a member of the Korean Academy of Science and Technology and the National Academy of Engineering of Korea. He is a Founding Member of the ACM/IEEE CPS Week and currently serving as a member of the steering committee for the IEEE RTCSA. He received the Outstanding Contribution Award from the Cyber Physical Systems Week in 2012. He has served on the Editorial Board of the ACM Transactions on Cyber Physical Systems, the IEEE Transactions on Computers, the IEEE Transactions on Parallel and Distributed Systems, and Real Time Systems journal.

JEONGGIL KO (Senior Member, IEEE) received the B.Eng. degree in computer science and engineering from Korea University in 2007 and the Ph.D. degree in computer science from Johns Hopkins University in 2012. He is an Assistant Professor with the School of Integrated Technology, Yonsei University. At Johns Hopkins University, he was a member of the Hopkins Inter Networking Research Group (HiNRG), led by Dr. A. Terzis. In 2010, he was at the Stanford Information Networking Group (SING), Stanford University, as a Visiting Researcher, with Dr. P. Levis. From June 2012 to August 2015, he was a Senior Researcher at the Electronics and Telecommunications Research Institute (ETRI), and he was an Assistant Professor at the Department of Software and Computer Engineering, Ajou University, from September 2015 to August 2019. His research interest includes developing embedded and mobile computing systems with ambient intelligence. He was a recipient of the Abel Wolman Fellowship awarded by the Whiting School of Engineering, Johns Hopkins University, in 2007.

---

**References**

[26] Cardiac Monitors, Heart Rate Meters, and Alarms, Standard ANSI/AAMI EC13-2002, 2002.

[27] Samsung Electronics. Samsung Health. Accessed: Oct. 11, 2020. [Online]. Available: https://www.samsung.com/global/galaxy/apps/samsung-health/

[28] S. Benedetto, C. Caldato, E. Bazzan, D. C. Greenwood, V. Pensabene, and P. Actis. “Assessment of the fitbit charge 2 for monitoring heart rate,” PloS ONE, vol. 13, no. 2, Feb. 2018, Art. no. e0192691.

[29] Tizen Developers, Ed. Monitoring Performance With Dynamic Analyzer. Accessed: Oct. 11, 2020. [Online]. Available: https://developer.tizen.org/development/tizen-studio/native-tools/debugging-your-app/dynamic-analyzer

---

**Acknowledgment**

This work was supported in part by the Ministry of Science and ICT, Korea, under the National Research Foundation, Grant No. NRF-2019R1A2C1087828.