Passive Vital Sign Monitoring via Facial Vibrations Leveraging AR/VR Headsets

Tianfang Zhang[1], Cong Shi[2], Payton Walker[3], Zhengkun Ye[4], Nitesh Saxena[3], Yan Wang[4], Yingying Chen[1]
[1] Rutgers University, [2] New Jersey Institute of Technology, [3] Texas A&M University, [4] Temple University
{tz203,yingche}@scarletmail.rutgers.edu,{cong.shi}@njit.edu,{prw0007,nsaxena}@tamu.edu,
{zhengkun.ye,y.wang}@temple.edu

ABSTRACT

Vital signs (e.g., breathing and heart rates) and personal identities are essential information for personalized medicine and healthcare. The popularity of augmented reality/virtual reality (AR/VR) provides an excellent opportunity for enabling long-term health monitoring in a broad range of scenarios, including virtual entertainment, education, and telemedicine. However, commercial-off-the-shelf AR/VR devices do not have dedicated biosensors for providing vital signs and personal identities. In this work, we propose a novel framework that can generate fine-grained vital sign signals and other personalized health information of an AR/VR user through passive sensing on AR/VR devices. In particular, we find that the user's minute facial vibrations induced by breathing and heart beating can impact the readily available motion sensors on AR/VR headsets, which encode rich vital sign patterns and unique biometrics. The proposed framework further estimates the breathing and heartbeat rates, detects the gender and identity, and derives the body fat percentage of the user. To mitigate the impacts of body movement, we design an adaptive filtering scheme to cancel the spontaneous and non-spontaneous motion artifacts. We also develop unique facial vibration features and deep learning techniques to facilitate vital sign signal reconstruction and user identification. Extensive experiments demonstrate that our framework can achieve a low error of vital sign signal reconstruction and rate measurement, along with 95.51% and 93.33% accuracy on identity and gender recognition.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*.

KEYWORDS

Health Monitoring; Facial Vibrations; AR/VR Headsets

ACM Reference Format:
Tianfang Zhang, Cong Shi, Payton Walker, Zhengkun Ye, Nitesh Saxena, Yan Wang, Yingying Chen. 2023. Passive Vital Sign Monitoring via Facial Vibrations Leveraging AR/VR Headsets. In *The 21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23), June 18–22, 2023, Helsinki, Finland*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3581791.3596848

MobiSys '23, June 18–22, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0110-8/23/06…$15.00
https://doi.org/10.1145/3581791.3596848

1 INTRODUCTION

The adoption of augmented reality/virtual reality (AR/VR) has risen dramatically over the past few years across various sectors, including immersive gaming, social events, education, and tourism. The global market size of AR/VR is expected to reach 43.01 billion USD at an annual growth rate of 27.5% in 2028 [5]. The emerging use of AR/VR creates an excellent opportunity to promote pervasive health monitoring since most AR/VR devices are already equipped with decent sensing and computing technology and will interact with users for a long time. In this work, we want to explore innovative technologies enabling fine-grained health monitoring (e.g., vital signs and user identities) via AR/VR devices.

Importance of Health Monitoring in AR/VR. Enabling fine-grained health monitoring in AR/VR adds a complementary factor to remote healthcare monitoring to the general public. On the one hand, it can provide real-time health information required in many virtual healthcare and education applications. For instance, a VR doctor can continuously monitor a patient's vital signs during and after the telemedicine session at home, which helps the doctor to make more precise diagnoses [17]. A virtual educator can exploit the real-time vital signs of students to improve their learning efficiency (e.g., reducing distraction) in a virtual classroom [14]. On the other hand, as people spend increasing time in cyberspace (e.g., Metaverse [4]), continuous exposure to virtual environments requires high concentration on the mind. Such effort may significantly increase the visual and psychological burden and cause various health issues (e.g., anxiety, hypertension, and sleep disorders) [6, 20, 44], especially for the younger generations who are undergoing vision and brain development. Thus, monitoring vital signs and providing timely health recommendations to users when using AR/VR devices is essential.

Existing health monitoring solutions mostly rely on medical instruments [23, 44] and dedicated biosensors (e.g., photoplethysmography (PPG) sensors and respiration monitoring belts) [26, 31]. While the current generation of AR/VR devices has various built-in sensors (e.g., motion sensors, position sensors, and front cameras), they are designed for immersive human-computer interactions in virtual environments. Compared to biosensors, the sensors readily available on AR/VR devices cannot provide health information directly. Note that PPG sensors are equipped in wearable devices, but they are unlikely to be integrated into the current generation of AR/VR devices, especially for low-cost headsets using a smartphone

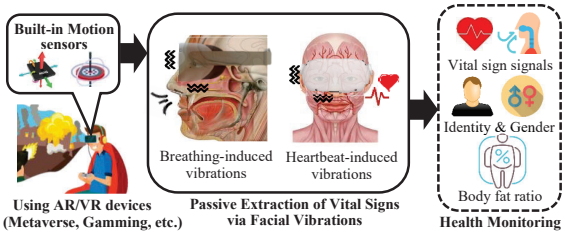


Figure 1: Illustration of the proposed health status monitoring framework leveraging facial vibrations captured by built-in motion sensors of AR/VR headsets.

as the screen and processing unit. In addition, PPG sensors only obtain heartbeat signals from the wrist and do not capture breathing patterns. All these lead to a renewed search for an integrated health monitoring solution that captures both types of vital sign signals using sensors already embedded in AR/VR devices.

Proposed approach based on vital-sign-induced facial vibrations. In this work, we find that subtle facial vibrations induced by breathing and heartbeat can impact built-in AR/VR motion sensors. The key insight is that the conductive vibrations going through the user's cranial bones and facial muscles will vibrate face-mounted AR/VR headsets. The muscle contract and relaxation related to the inhaling and exhaling in the breathing cycles inject physiological traits (e.g., facial geometry, facial structure) into the motion sensor readings. Furthermore, as the facial arteries normally contact the AR/VR headset during usage, the AR/VR headset picks up minute vibrations induced by the fluctuation of blood flow in the human's face, which encodes rich physiological traits of the user's cardiovascular system. It is similar to taking radial pulses on the wrists using fingers. Inspired by these findings, we propose to perform passive health monitoring by extracting vital-sign-induced facial vibrations. Through the development of a novel deep learning technique, our framework continuously reconstructs vital sign signals that bear similar signal quality as those captured by dedicated medical instruments: head-mounted Arduino PPG sensors and respiration monitoring belts. Based on the fine-grained waveforms of breathing and heart beating, the framework then estimates the breathing and heartbeat rates, detects the gender and identity, and even derives the user's body fat percentage. Our framework can serve as a general building block for many AR/VR healthcare applications, such as well-being monitoring, symptom diagonal, healthcare recommendations, etc. The concept of the proposed vital-sign-induced vibration-based personalized health monitoring framework is illustrated in Figure 1.

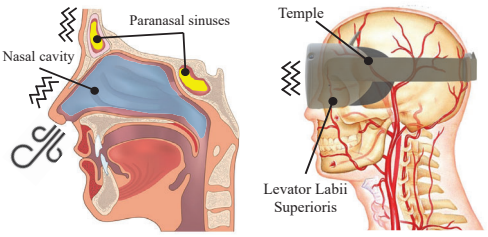
Difference from Existing Approaches. Our health monitoring framework manifests significant differences from existing health monitoring methods (e.g., vital sign monitoring) using motion sensors and various mobile and wearable devices. Several existing works try to attach a smartphone to the users' chests to measure the vital-sign-related chest movements with the smartphone's accelerometer [29, 49]. These approaches, however, constrain the user's posture to lying down and interfere with other normal activities. Differently, our framework can passively obtain vital sign while the user is enjoying AR/VR applications. Developing a passive solution is far more challenging compared to the smartphone-based approach as it requires capturing minute facial vibrations caused by blood flow and nasal cavity motions. Our work is also different from existing heartbeat monitoring solutions based on the PPG sensors of wristbands/smartwatches, which cannot capture users' breathing patterns. For our framework, it integrates breathing and heart beating monitoring, gender recognition and user identification capability into a single AR/VR headset without installing/wearing additional sensors. In contrast, traditional and medical approaches require users to attach different devices to multiple parts of their body (e.g., respiration belts, wearable devices, PPG sensors) [18, 27, 42, 47]. It can be used in complement with existing mobile healthcare applications [8, 53]. For instance, in AR-enabled telehealth/telemedicine appointments, patients can put on AR/VR headsets instead of multiple dedicated sensors to provide necessary health information and verify their identity with the doctor at home. In addition, our framework is passive as it does not require any active user inputs (e.g., typing names and passwords). It is thus friendly to users with disabilities or inconvenient to interact with AR/VR devices.

Our contributions are three-fold:

(1) ***A Novel Health Monitoring Framework for AR/VR Users:*** We develop a novel health monitoring framework that exploits facial vibrations to derive users' personal health-related information, including breathing patterns, heartbeat signals, gender, identity, and body fat percentage. It is the first work demonstrating that built-in motion sensors of commercial AR/VR headsets can capture subtle vital-sign-induced facial vibrations carrying biometrics and vital signs. The designed framework can serve as a building block for various healthcare applications in future AR/VR paradigms.

(2) ***Fine-grained Physiological Traits Extraction:*** To enhance the framework robustness under various motions in AR/VR application scenarios, we design an adaptive algorithm to remove both spontaneous and non-spontaneous body movements. A sensor fusion scheme is further developed to constructively combine sensor readings, meanwhile mitigating the influence of different wearing styles. We further design a deep-learning-based technique to reconstruct precise breathing and heartbeat signals, with signal quality approximating dedicated medical instruments, thereby enabling fine-grained health monitoring. Extensive experiments on three AR/VR headsets under various settings show that our framework can achieve accurate breathing/heartbeat signal reconstruction with low root mean square errors.

(3) ***Personalized Health Information Derivation:*** We design a user identification scheme solely based on facial vibrations, which personalizes the vital sign data for accurate diagnoses of health conditions. Our scheme extracts representative features to characterize the unique facial characteristics of individual users (e.g., facial shape, geometries, fat content). Our scheme also supports gender detection, which allows applications to provide gender-specific health recommendations. We further demonstrate that a correlation exists between body fat percentage and facial vibrations by designing a deep regression-based scheme. Extensive experiments show that our scheme can achieve over 94% accuracy in both user identification and gender detection and less than 13% error rate in body fat percentage estimation.



(a) Breathing-induced vibrations (b) Heartbeat-induced vibrations

Figure 2: Facial structures involved during human's breathing and heartbeat.

2 PRELIMINARIES

2.1 Potential Applications of Our Work

We envision that the precise breathing and heartbeat signals extracted from facial vibrations can be correlated with a broad array of physiological and psychological statuses, which facilitates many existing and emerging applications. We list a few potential applications of our system as follows:

Well-being Monitoring. Continuous and remote well-being monitoring, especially vital sign tracking along with gender recognition, facilitate the timely detection of diseases and the identification of medical parameters. For example, a doctor can use our framework as an AR/VR medical service to obtain gender and vital sign information, which helps to identify some gender-specific diseases, such as breast cancer, osteoporosis and color blindness. Furthermore, the body fat percentage information provided by our framework offers additional information for detecting diseases related to overweight and obesity, including diabetes, hypertension, fatty liver, etc. In general, the proposed framework can enhance the effectiveness of telehealth/telemedicine, which solely relies on voice and video nowadays [25, 32]. It also allows patients to measure their vital signs for the long term at home using low-cost AR/VR devices, which provides more valuable and comprehensive health data to doctors/medical specialists.

Virtual Education. AR/VR is bringing fundamental changes to education by creating opportunities for virtual narrative or demonstration [59], interactive training [57], and 3D art design [33]. Particularly, our framework could be incorporated into attention-monitoring apps for virtual education. During virtual learning, students interact with the instructor and virtual learning objects in a virtual environment. The students could be inactivity due to tiredness or distraction after long-term concentration. With our framework, the instructors will be able to detect and locate inactive students by analyzing the breathing and heartbeat signals, which helps to quantify the students' learning efficiency. Such knowledge can help educators dynamically adjust their strategies and contents to maximize teaching performance.

Cybersickness Detection/Mitigation. There has been active research on mitigating cybersickness in the AR/VR community. Cybersickness frequently occurs when people are exposed to virtual environments, where the virtual motions conflict with the expected ones from their brains. Existing approaches of detecting cybersickness rely on questionnaires [20, 50] or heartbeat variation monitoring using dedicated instruments (e.g., PPG [23, 44] and ECG sensors [13, 39]). Such approaches are time-consuming and intrusive. Differently, our framework extracts vital sign signals based on motion sensor data, which facilitates detecting cybersickness without installing/wearing additional sensors. It may also benefit the research on cybersickness by collecting fine-grained vital sign data from a large group of AR/VR users (e.g., supporting HealthData.gov programs [24]).

2.2 Vital-Sign-Induced Vibrations

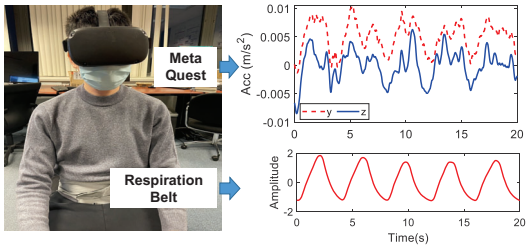
There are two types of vital-sign-induced vibrations that can be captured by AR/VR headsets: breathing- and heartbeat-induced facial vibrations.

Breathing-Induced Vibrations. Breathing is closely related to the contraction and relaxation of nasal cavities located in the upper respiratory tract of humans' heads. The inhaling (breathing in) and exhaling (breathing out) contract and relax the nasal cavities, resulting in movements of facial muscles and tissues in a synchronous fashion. In addition, the volume variations of paranasal sinuses, which are in direct connection with the nasal cavities, also generate vibrations correlated with human breathing. These vibrations of nasal cavities and paranasal sinuses carry unique breathing characteristics (e.g., amount of airflow, and breathing duration and magnitude) and physiological traits (e.g., paranasal sinuses and nasal muscle properties). During the usage of AR/VR headsets, users' nasal cavities and paranasal sinuses are in direct contact with the head-mounted devices, thereby impacting the built-in AR/VR motion sensors. The anatomy of nasal cavities, paranasal sinuses, and their relative positions to the AR/VR headset are illustrated in Figure 2(a).

Heartbeat-Induced Vibrations. The human heart pumps blood through alternative contraction and relaxation, forming periodic patterns of blood flow in vessels, including the arteries and veins across the human face. A typical heartbeat cycle includes four major phases (i.e., atrial systole, isovolumetric contraction, ventricular ejection, and isovolumetric relaxation), and they induce a series of systolic and diastolic points in the blood flow patterns [10]. When wearing an AR/VR headset, the arteries/veins close to the levator labii superioris [11] of the user intimately contact the headset, thereby encoding the blood flow patterns as minute vibrations to the headset, including the systolic and diastolic points. In addition, the headset is in direct contact with the temple of the user's head, which is surrounded by rich arteries and produces facial vibrations associated with heart beating. We show the blood vessel distribution on the human's face and their relative locations with respect to the AR/VR headset in Figure 2(b).

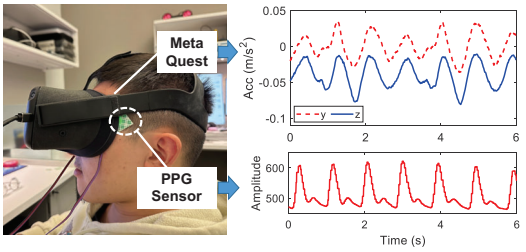
2.3 Capturing Breathing and Heart Beating via Facial Vibrations

Motion sensors (i.e., accelerometer and gyroscope) are built into most AR/VR headsets for head motion tracking, either controlling the AR/VR field of view or reconstructing the user's head movements in the virtual world. Besides measuring acceleration and angular velocity, existing works have also proved that these sensors are able to capture vibration signals [9, 52], thus also enabling them to capture the aforementioned vital-sign-induced facial vibrations.



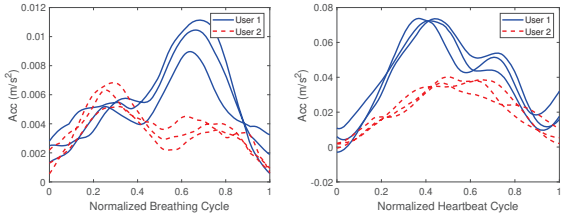
(a) Human subject (b) Accelerometer and respiration belt readings

Figure 3: Breathing-induced facial vibrations on accelerometer (y-axis and z-axis) readings of Meta Quest and NeuLog NUL-236 Respiration Monitor Belt.



(a) Human subject (b) Accelerometer and PPG sensor readings

Figure 4: Heartbeat-induced facial vibrations on accelerometer (y-axis and z-axis) readings of Meta Quest and PPG readings on a real user.



(a) Breathing patterns of 2 users (b) Heartbeat patterns of 2 users

Figure 5: Comparison of facial vibrations on accelerometer (Z-axis) of two different users on Meta Quest.

To examine the effects of breathing-induced facial vibrations, we conduct a preliminary study by asking a user to wear a commercial AR/VR headset (i.e., Meta Quest [38]), sit on a chair, and breathe normally. A NeuLog NUL-236 respiration monitor belt is worn on the user's belly as shown in Figure 3(a) to collect the pressure associated with the expansion and contraction of the belly as ground truth. The sampling rates of the motion sensors and the respiration monitor belt are $1000Hz$ and $50Hz$, respectively. We compare the accelerometer readings (i.e., processed with a band-pass filter of $0.1Hz \sim 0.5Hz$) and pressure measurements as shown in Figure 3(b). The two types of readings exhibit similar waveforms and periodicity, showing the sensitivity of the accelerometer to breathing-induced vibrations. We conduct a similar experiment by asking the user to wear a Meta Quest and a head-mounted Arduino PPG pulse sensor as shown in Figure 4(a). Figure 4(b) illustrates a similar periodicity in both the motion sensor readings (i.e., processed with a band-pass filter of $0.8Hz \sim 3.0Hz$) and the PPG measurements. An interesting finding is that the systolic and diastolic points can also be captured in motion sensor readings, validating the correlation between the heartbeat patterns and the facial vibrations.

We further conduct an experiment to study the potential of capturing unique breathing and heartbeat properties based on facial vibrations. Figure 5(a) and (b) shows the breathing- and heartbeat-induced vibrations of two different users. As a result of these studies, we found that two users had different morphological properties (e.g., breathing patterns, systolic and diastolic peaks) in both types of facial vibrations, indicating that these properties can be used to differentiate users. We will demonstrate that facial vibrations can be leveraged for gender detection and body fat percentage estimation in Section 6 and 7.

3 FRAMEWORK DESIGN

3.1 Challenges

Significant Impacts of Body Motions. During practical usage of AR/VR headsets, a user interacts with the virtual world via different scales of body motions, such as moving the controllers and rotating his/her heads. Such motions significantly interfere with the facial vibrations, distorting the embedded vital sign patterns. T have orders of magnitude higher than those of minute vibrations induced by vital signs, inducing a significant amount of noise. We need to design a scheme to mitigate the impacts of such distortions.

Extracting Precise Breathing and Heartbeat Signals. Vital sign patterns, especially some manifestations (e.g., systolic and diastolic points), are usually utilized for health status detection and disease diagnosis [15, 46, 62]. However, motion sensors have low sensitivity to such vital sign patterns/manifestations due to their minute magnitude, making them sometimes "invisible" in raw motion sensor readings. To facilitate healthcare applications, it is essential to reconstruct waveforms of the respiration and heartbeat patterns.

Biomarker and Biometric Derivation for Healthcare Monitoring. The framework based should have the capability to support various healthcare applications, which require various kinds of biomarkers (e.g., breathing rate, heartbeat rates) and biometrics (e.g., user-specific characteristics). Thus, the framework needs to extract various types of features that comprehensively characterize the user's physiological status.

3.2 Framework Overview

To address the aforementioned challenges, we design a health monitoring framework leveraging vital-sign-induced facial vibrations as illustrated in Figure 6. The key idea of our framework is to reconstruct precise vital sign signals from motion sensor readings, which are passively collected during the usage of AR/VR devices. We further showcase the feasibility of a suite of applications (i.e., breathing & heartbeat rate estimation, user identification, gender detection, and body fat percentage estimation).

Facial Vibration Extraction. Our frameworks take time-series accelerometer and gyroscope readings as input. When the data are being collected, facial vibrations are mixed with body motions (e.g., head tremors, movements for VR interactions), which degrades the sensitivity to vital sign patterns. To remove motion artifacts, we design a *Motion Artifact Cancellation* module to adaptively cancel these distortions leveraging neighboring segments without being impacted by the motions. To combat the orientation dynamics of

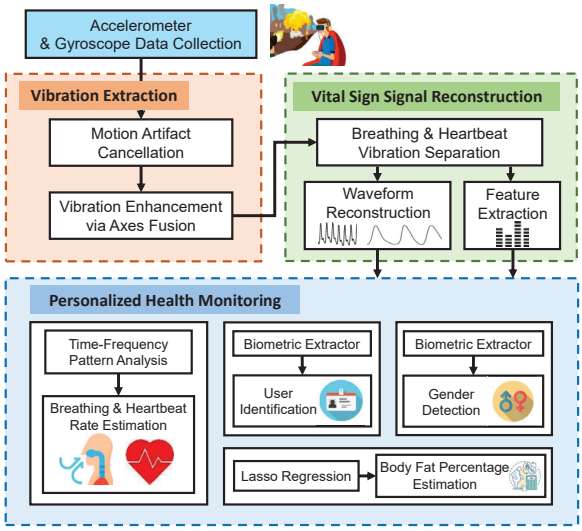
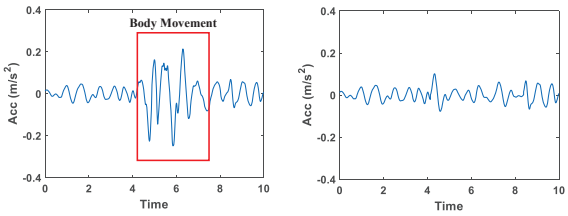


Figure 6: Overview of the proposed personalized health monitoring framework.

the headset (e.g., when users rotate their heads to view moving virtual objects), we design a *Dynamic Axes Fusion* approach to select and combine sensor readings from three axes on the accelerometer and gyroscope, respectively.

Vital Sign Signal Reconstruction. Next, our framework separates facial vibrations of breathing and heart beating and reconstructs precise vital sign signals, with signal quality approximating medical sensors. As breathing and heart beating normally have distinctive frequency ranges, we use two band-pass filters to isolate the breathing and heartbeat vibrations, respectively. For each type of vibration, we compute spectrograms as features, which reveal small-step frequency variations of breathing/heart beating. Then, a deep learning model is designed to reconstruct signals approximating pressure measurements of breathing and PPG-like patterns of heart beating. Particularly, we explore bidirectional long short-term memory (LSTM) units with an attention mechanism in our model design, which facilitates learning a reliable mapping relationship between facial vibrations and the vital sign signals across human subjects. A model built on a group of people can be directly applied to new users without additional training data. Besides signal reconstruction, our framework also extracts a set of representative features of facial vibrations to support health applications.

Personalized Health Monitoring. We showcase four types of applications built upon the reconstructed vital sign signals and facial vibration features to support health monitoring: *Breathing & heartbeat Rate Estimation:* Our framework estimates the breathing and heartbeat rates by detecting the prominent frequencies of the reconstructed breathing and heartbeat sign waveforms; *User Identification:* The framework personalizes the health monitoring process through identifying users. Particularly, we design a model based on a convolution neural network (CNN) to derive representations correlated with users' unique biometrics (e.g., facial geometry, facial structure) for user identification; *Gender Detection:* As pathology conditions are normally correlated with gender [3, 30], our framework also performs gender detection. We build a lightweight classifier based on a support vector machine (SVM) based on the



(a) Readings before cancellation (b) Readings after cancellation

Figure 7: Comparison of heartbeat-induced facial vibrations on accelerometer (z-axis) before and after body movement cancellation.

abstractions derived from the waveform reconstruction model to detect the user's gender; *Body Fat Percentage Estimation:* Prior research shows the correlation relationship between breathing/heartbeat signals (e.g., PPG, ECG) and body fat percentage [7, 56]. We demonstrate such a correlation still exists when using the reconstructed waveform of our framework, by designing a lasso-regression-based approach to estimate the body fat percentage.

4 VITAL-SIGN-INDUCED FACIAL VIBRATION EXTRACTION

4.1 Motion Artifact Cancellation

In practical scenarios, AR/VR users interact with the headsets through different scales of body motions, especially non-continuous movements such as intermittent walking and head rotations. To achieve reliable facial vibration extraction, we need to decouple the motion sensor readings associated with vital signs and those induced by such motions. Particularly, we design a real-time scheme based on dynamic filtering [48] to remove the artifacts of non-continuous motions. Our scheme first computes time-series Short-Time Energy (STE) upon motion sensor readings sampled with a sliding time window, which reveals the regions with body movements. The scheme then applies an adaptive filter on readings of these regions to dynamically remove the motion artifacts.

Energy-based motion detection. Our scheme calculates STE using a time window of length N as: $STE(t) = \sum_{n=t}^{t+N} x^2(n)$, where $x(\cdot)$ can be either time-series accelerometer or gyroscope readings. We empirically choose a window size $N = 3000$ (data of 3 seconds). A threshold $\tau = 0.2$ is used to detect starting and ending points in the motion sensor readings, where the motion artifacts reside in, while skipping the regions of breathing and heart beating.

Adaptive Filter. Our scheme then compute a weight vector w for each region with motions. We model this process as an optimization problem. Specifically, for the readings of each detected region, the scheme finds the nearest data segment of the same length that is not contaminated with body motions (i.e., STE is lower than threshold τ). The segment is then used as the reference signal to compute a w rendering similar morphology characteristics in the contaminated sensor readings. The objective function is defined as:

$$\begin{aligned} &\arg\min_{w} \sum_{t=1}^{T} \left|r(t) - y(t)\right|^2, t \in [1, T], \\ &s.t.\ e(t) = \left|r(t) - y(t)\right|^2, \\ &\quad w(t) = \alpha w(t) + \mu e(t) x(t),\ \ y(t) = w(t) x(t), \end{aligned} \tag{1}$$

where y, r, and e represent vectors of the denoised vibration signals, the reference signals, and the estimation error respectively. T denotes the length of the vectors (i.e., the number of data points). α is the leakage parameter of the filter and μ denotes the step size for parameter updating. We conduct an experiment by asking a user to wear Meta Quest and browse a webpage using a controller. A comparison before and after cancellation is shown in Figure 7. After filtering, the motion artifacts are mostly removed and replaced by vital-sign-induced responses. The results confirms the effectiveness of our scheme on removing non-continuous body motions.

4.2 Vibration Enhancement via Axes Fusion

Due to the variations of wearing positions/styles of the headset, the three axes of the accelerometer and gyroscope exhibit differences in their sensitivity to vital-sign-induced facial vibrations across sessions. A subset of axes may show higher sensitivity than the other axes, depending on the orientation of the headset. To remove the impact of headset orientation, we design an axes-fusion scheme that dynamically selects of the three axes of motion sensors. We leverage the fact that the sensitive axes normally show higher magnitudes regarding periodic signals, which include both breathing and heartbeat patterns in AR/VR settings. To quantify the sensitivity, we apply Fisher's Kappa [40, 41] upon the vibration signals, which measure the periodicity strength on each axis of accelerometers and gyroscopes:

$$\begin{aligned} score\big(x(t)\big) &= \frac{max\Big(PSD\big(x(t)\big)\Big)}{\sum_{i=1}^{T}\frac{PSD\big(x(i)\big)}{T}}, t \in [1,T], \\ s.t.\ PSD\big(x(t)\big) &= \frac{1}{T}10log_{10}\Big(S\big(x(t)\big)\Big)^2, \\ S\big(x(t)\big) &= abs\Big(\underset{0.8\sim3.0Hz}{FFT}\big(x(t)\big)\Big), \end{aligned} \tag{2}$$

where $PSD(\cdot)$ represents the power spectral density of the signal $x(t)$. Bandpass filters with the frequency range of heartbeat ($0.8Hz \sim 3.0Hz$) [2] along with fast Fourier transform (FFT) are applied to motion sensor readings to compute the power $S(x(t))$ within the frequency range. $S(x(t))$ is further utilized to compute the PSD and Fisher's Kappa scores for axes fusion. A higher score refers that the sensor and its corresponding axis is more sensitive to vital-sign-induced vibrations for this user. Particularly, the value of Fisher's Kappa of reading streams is computed from all sensor-axis combinations and we choose the first three combinations with the top Fisher's Kappa per user. Note that we only quantify the sensitivity to heart beating during the axes fusion. This is because we empirically find that breathing induces larger vibrations than heart beating, and the combined measurements using heartbeat are also sensitive to breathing.

5 RECONSTRUCTION OF BREATHING AND HEARTBEAT SIGNALS

5.1 Signal Separation of Facial Vibrations

The facial vibrations of breathing and heart beating are mixed in the motion sensor readings. As these two types of vital signs have distinctive frequency ranges, we use two band-pass filters with different cut-off frequencies to separate the vibrations of breathing and heart beating. Periods of human breathing are normally within $12 \sim 16$ repeats per minutes [1] (i.e., $0.1Hz \sim 0.5Hz$), while heart beating ranges $60 \sim 100$ beats per minutes [2] (i.e., $0.8Hz \sim 3.0Hz$). Thus, we use these two sets of frequencies as the cut-off frequencies to decouple the breathing and heartbeat vibrations.

5.2 Waveform Reconstruction Model

Model Overview. As we capture facial vibrations in terms of vital sign patterns using low-grade built-in motion sensors, the morphological characteristics are more vulnerable to hardware noises compared to those captured by medical instruments. For example, the systolic and diastolic points are sometimes "missing" in motion sensor readings due to their weak millimeter-level magnitude. To mitigate such hardware noises and reconstruct precise vital sign patterns, we have developed a deep-learning-based model. Since vital sign signals indicate strong periodical and sequential characteristics, we build a deep reconstruction model based on long-short-term-memory (LSTM) to expose these features. Additionally, to capture internal dependencies within each data segment, we incorporate a self-attention mechanism [60] to establish the correlations between our captured facial vibrations and ground truth respiration and heartbeat signals. Through learning the correlation along with temporal dependency, our model built on data of some people can be directly applied to reconstruct vital sign signals of new users, without the need of collecting data of new users from medical instruments. We show the architecture of our reconstruction model in Figure 8. The model's input consists of 3 channels, $C1$, $C2$ and $C3$, corresponding to 3 pre-selected combinations of motion sensors and axes. For each channel, it takes a set of segments $\{x_1, ..., x_N\}$ (i.e., breathing or heartbeat vibrations) as input, which contains N segments of vibrations of length T (i.e., number of data points in 3 seconds). The ground truth set $\{p_1, ..., p_N\}$ is also collected using the respiration monitoring belt and PPG sensor. For each segment in $\{x\}$ and $\{p\}$, we use Short-Time Fourier Transform (STFT) to compute spectrograms that reveal small-step frequency variations of breathing/heart beating. A temporal feature extractor $D(\cdot)$ is then applied to convert the spectrograms of x into a set of vital-sign representations. Then, we train a waveform reconstructor $G(\cdot)$ to learn to map the vibration spectrograms into derived from P, making it capable of reconstructing signals approximating those collected with medical instruments.

Objective. For the forward propagation of the reconstruction model, a segment of breathing/heartbeat vibration x is fed to a temporal feature extractor $D(\cdot)$ and a waveform reconstructor $G(\cdot)$, with the signal y as reconstruction output. Given the training data $\{x\}$ and ground truth $\{p\}$, we optimize both $D(\cdot)$ and $G(\cdot)$ based on a 2D mean square error (MSE), which quantifies the difference between the spectrograms of the reconstructed signals $\{y\}$ and those of $\{p\}$. The 2D MSE loss L is defined as:

$$\begin{aligned} L(y_i, p_i) &= \frac{1}{T}\cdot\frac{1}{F}\sum_{t=1}^{T}\sum_{f=1}^{F}\Big(s(y_i,t,f) - s(p_i,t,f)\Big)^2, \\ s.t.\ s(y_i,t,f) &= G\Big(D\big(s(x_i,t,f)\big)\Big), \end{aligned} \tag{3}$$

where $s(y,t,f)$ represents the spectrogram, with t and f representing the time and frequency indices, respectively. During training, features in both time and frequency dimensions are reconstructed to match the ground truth. We define the overall objective of the

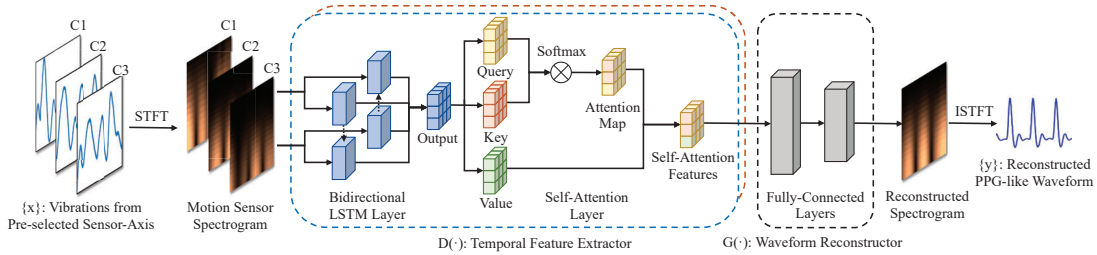
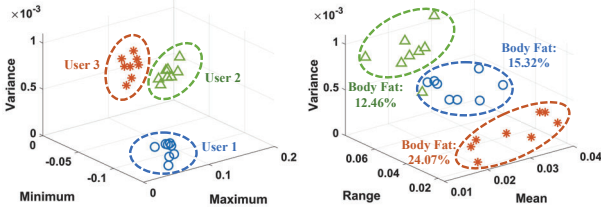


Figure 8: Deep learning architecture for vital sign signal reconstruction.



(a) Features on identity (b) Features on body fat

Figure 9: Time domain statistic features for distinguishing users and different body fat percentages.

optimization process as: $\arg\min_{D,G} \sum_{i=1}^{N} L(y_i, p_i)$, where D and G denote the weights of the extractor $D(\cdot)$ and the reconstructor $G(\cdot)$, respectively.

Temporal Feature Extractor. Vital sign signals normally exhibit unique temporal patterns, with fiducial points appear one after another. To capture such temporal patterns, we design a network based on bidirectional long short-term memory (LSTM) units with self-attention mechanism [51] as shown in Figure 8. The ability of self-attention allows the model to automatically focus on critical temporal regions for reconstructing the vital sign signals. We set the filter numbers of the two bidirectional LSTM layers as 1024 and 512, respectively. For two self-attention layers, we make the query (Q), key (K), and value (V) equal to the output of their connected LSTM layers with ReLU as activation functions.

Waveform Reconstructor. Based on features from $D(\cdot)$, we develop a reconstructor to reconstruct spectrograms of vital sign waveforms. The reconstructor consists of 2 fully-connected layers with ReLU as the activation function. The numbers of output neurons are 512 and 256 in two layers.

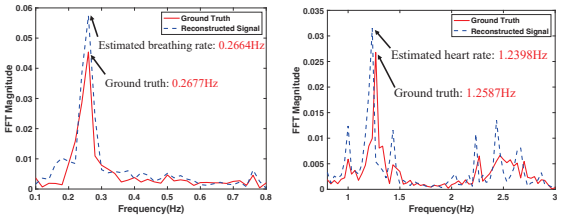
5.3 Representative Feature Extraction

Besides reconstructing vital sign signals, we also extract features to elicit biomarkers and biometrics embedded in facial vibrations, which are physiological characteristics beyond breathing and heartbeat patterns.

Time-Domain Statistic Features. In Section 2.3, we conduct preliminary experiments and observe that the morphological patterns of respiration and heartbeat differ significantly among different users. We believe the morphological characteristics of vital signs capture unique face structures and muscle properties of each individual, which result in distinct time-series vibration patterns. These morphological patterns can be effectively represented with time-domain statistic features. For example, the maximum value of a signal segment reflects the strongest response of facial vibrations for a specific user, while the range of a signal segment can describe the amount of air inhalation or blood flow associated with a person's breathing and heartbeat. By leveraging these statistic features, we are able to accurately capture individual differences in vital sign patterns for different healthcare applications. To extract the time-domain characteristics of facial vibration signals, we first split the signals into short segments of 3 seconds each, which are very short-time periods in AR/VR scenarios. Subsequently, we apply a sliding window with 256 points on our segmented short frames of breathing/heartbeat vibrations and extract 13 features: maximum, minimum, range, mean, variance, root mean square, median, interquartile range, mean crossing rate, skewness, kurtosis, entropy, and power. The time-domain features are utilized for user identification and body fat percentage estimation, which we will introduce in Section 6.1 and Section 6.3. We example features variance, minimum, and maximum of three users in Figure 9 (a). We can find that the feature clusters are differentiable among the users. Similarly, we observe in Figure 9 (b) that features form unique clusters for different body fat percentages.

Frequency-domain Features. The breathing and heartbeat vibrations also introduce vibration patterns in the frequency domain, which exhibit user-specific patterns. For instance, the strongest responses vary among different users and are associated with the cycles respiration/heartbeat and corresponding facial vibration intensity. Moreover, some users exhibit more pronounced harmonics of heartbeat vibrations, which may be attributed to differences in their face structures and components. To capture these user-specific characteristics from the frequency domain, we extract frequency-domain features by applying Short Time Fourier Transform (STFT) on breathing/heartbeat signal segments with 3 seconds lasting, where the vibration responses at different frequencies can be revealed. We demonstrate how we leverage the features for user identification in Section 6.1.

Gender-Related Hidden Features. Prior works [34, 37] have shown the correlations between gender and heartbeat patterns captured by PPG sensors in a large population. However, these studies primarily focus on differentiating genders by analyzing systolic and diastolic features, which could be distorted in the presence of noise in motion sensor readings.Thus, it is more challenging to extract gender-related features directly from vibration signals. In this work, we proposed an innovative approach to extract gender-related hidden features for gender recognition. Specifically, we reuse the output of the first fully-connected layer (1×1024) in the waveform reconstructor, as it reserves features from both motion sensor readings and PPG waveforms. We find that these hidden



(a) Response of breathing (b) Response of heartbeat

Figure 10: Frequency response distribution of reconstructed breathing and heartbeat waveforms.

features well characterize the gender of users, with the techniques introduced in Section 6.2.

6 PERSONALIZED HEALTH MONITORING

6.1 Breathing and Heart Rates Estimation

To estimate the breathing and heart beating rates, we apply time-frequency analysis on the reconstructed breathing and heartbeat signals. Specifically, Hann Window is utilized to reveal signal periodicity and mitigate side lobe artifacts. Next, we apply FFT upon the reconstructed breathing and heartbeat signals to reveal their frequency responses. Finally, a peak selection algorithm is adopted within the range of $0.1Hz \sim 0.5Hz$ and $0.8Hz \sim 3.0Hz$, which are the common frequency ranges of human breathing and heart beating, to respectively detect breathing and heartbeat frequencies. Comparisons of the frequency responses between the reconstructed vital sign frequency responses and the ground truth are shown in Figure 10(a) and Figure 10(b). We observe that our framework can accurately detect vital sign rates based on the peak-detection-based approach.

6.2 User Identification

To realize user identification, we design a deep-learning model consisting of a representation extractor and a deep-learning-based classifier to identify users. It takes both time domain statistic features and frequency spectrogram features as input. As features of vital-sign-induced facial vibrations imply different properties in time and frequency domains, we develop two separate CNNs to extract the corresponding deep representations. These representations are then concatenated and fed into the classifier. Finally, a classifier with two fully-connected layers and a softmax layer is used for identity prediction (e.g., user's identity). For training the representation extractor and the identity classifier, we apply categorical cross-entropy loss as the loss function.

6.3 Gender Detection

We design a light-weighted classifier utilizing gender-related hidden features to achieve gender recognition. Specifically, we utilize a support vector machine (SVM) with a radial basis function (RBF) kernel. Based on hidden features with gender labels, the model learns to identify support vectors (high-dimension boundaries) to differentiate male and female features. We empirically find that the built support vectors based on a group of users can be applied to new users without new training data.

(a) Meta Quest (b) HTC Vive Pro Eye (c) Google Cardboard

Figure 11: Commodity AR/VR headsets involved in the experiments.

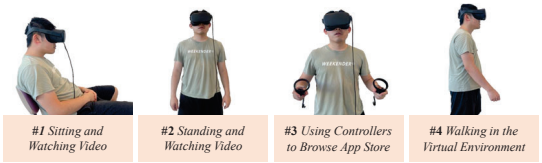


Figure 12: Illustration of four different scenarios.

6.4 Body Fat Percentage Estimation

Previous studies [54, 58] have demonstrated that breathing and heartbeat patterns can be roughly correlated with the body fat percentage. For instance, the mechanical effects of obesity will cause airway narrowing and closure, thus increasing resistance on human's respiratory system [16]. High body fat percentage will also alter users' cardiovascular systems by clogging the arteries with fat, blocking the blood flow and injecting distinctive traits into the heartbeat waveforms [43], thus influence the patterns of facial vibrations. Based on these insights, we develop a regression model to correlate users' body fat percentages with facial vibration features. Specifically, we choose *Lasso Regression Model* to make regression between time domain statistic features of facial vibrations and ground truth body fat percentage, which is built based on the least squares method with the Lasso constraint as below:

$$\sum_{i=1}^{M}(y_i - \hat{y}_i) = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{i=1}^{M} |w_j|, \quad (4)$$

where x, y, and w represent the time-domain statistic features, estimation results, and regression weights, respectively. M and p represent the number of data samples and characteristics for the regression model. λ works as the weight of the penalty term, which helps to control how many variants we want to keep. Empirically, we set it to be 0.5 in this task.

7 EVALUATION

7.1 Experimental Setup

AR/VR Headsets. We evaluate our framework on two standalone headsets (Meta Quest [38] and HTC Vive Pro Eye [61]) and one cardboard headset (Google Cardboard [12] with a Samsung Galaxy S6 smartphone). The sampling rate of motion sensors on Meta Quest, HTC Vive Pro Eye, and Samsung Galaxy S6 for our personalized health monitoring framework are $1000Hz$, $1000Hz$, and $203Hz$, which are the highest and most stable sample rates the headsets can achieve.

AR/VR Scenarios. We evaluate our framework in practical AR/VR scenarios involving different scales of body motions. Specifically, we select data from 4 common and representative scenarios as illustrated in Figure 12. *1) Sitting and Watching a Demo Video:* the participants are asked to sit in a chair and watch a demo video for 1 minute. During the experiment, the participants only use the

Table 1: Performance of reconstructed vital sign waveforms (RW) under same-user and cross-user settings.

Meta Quest			
	Facial Vibration	RW (same-user)	RW (cross-user)
Breathing	8.13	0.14 (↓ 7.99)	1.73 (↓ 6.40)
Heartbeat	27.83	0.46 (↓ 27.37)	3.50 (↓ 24.33)
HTC Vive Pro Eye			
	Facial Vibration	RW (same-user)	RW (cross-user)
Breathing	9.11	0.32 (↓ 8.79)	2.10 (↓ 7.01)
Heartbeat	24.56	0.38 (↓ 24.18)	4.95 (↓ 19.61)
Google Cardboard			
	Facial Vibration	RW (same-user)	RW (cross-user)
Breathing	10.19	0.31 (↓ 9.88)	2.61 (↓ 7.58)
Heartbeat	30.24	0.79 (↓ 29.45)	5.72 (↓ 24.52)

head-mounted display (HMD) for interaction and remain in a static position. *2) Standing and Watching a Demo Video:* the participants are requested to stand and watch a demo video for 1 minute. In this scenario, participants may exhibit some non-spontaneous body movements (e.g., body shaking) compared to more static and stable sitting scenarios. *3) Using Controllers to Browse the App Store:* different from previous scenarios of watching demo videos, participants are requested to browse the application store using controllers in this scenario. During the experiment, the participants perform spontaneous arm and hand movements, which are similar with arm lifting and dropping. Although we involve spontaneous motions in this scenario, these movements could not directly induce significant fluctuations on vibration signals captured from motion sensors on the AR/VR headset. Nevertheless, this scenario is more dynamic and complicated compared to the scenarios of watching demo video in a more static and stable manner. *4) Walking in the Virtual Environment:* the participants are asked to walk inside the virtual environment. Specifically, we ask the participants to walk a distance of about 3 steps (about 2 meters within the virtual environment) straight within roughly 3 seconds. The action of walking introduces large scale spontaneous movements different from aforementioned scenarios and directly induce significant changes on AR/VR motion sensor readings since the acceleration dramatically changes at the beginning and end of user's walking. We believe this scenario is able to summarize most types of large-scale, spontaneous and even more complicated body motions, such as head shaking, as they usually manifest similar significant and sudden fluctuations of motion sensor readings different from static scenarios.

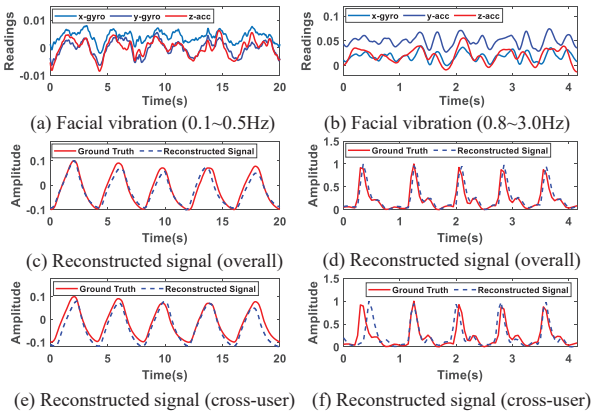
Data Collection. Our experiments involve 40 participants, the majority of whom are university students, with ages ranging from 20 to 44. Specifically, we have the same group of 25 users (17 males and 8 females) for both Meta Quest and HTC Vive Pro Eye. Another group of 15 participants (8 males and 7 females) is requested to wear the Google Cardboard headset with Samsung Galaxy S6 for data collection. Every participant is asked to wear the headset in four aforementioned scenarios. Since there is no controller for the Google Cardboard headset, participants are requested to perform arm movements with a similar scale as those scenarios in which participants are using Meta Quest or HTC Vive controllers. For all experiments, the participant wears a NeuLog NUL-236 Respiration Monitor Belt [36] and a NeuLog NUL-208 Photoplethysmography Monitor [35] to collect ground truth breathing and heartbeat signals with the collection of motion sensor readings. The duration of each experiment in each scenario is 1 minute. To evaluate the performance of body fat percentage estimation, we use a RENPHO Smart Body Fat Scale [45] to track the body fat percentage of 10 users for 1 month. Although the device only provides consumer level of body fat estimation, we adopt them as the ground truth to demonstrate our framework should be able to provide comparable results to specialized medical devices. The data collection has been approved by our university's IRB.

Evaluation Metrics. *1) Root Mean Square Error (RMSE).* We evaluate the performance of breathing/heartbeat pattern or waveform reconstruction module through computing RMSE between our reconstructed pattern of breathing or heartbeat cycle and patterns collected from specialized medical sensors (e.g., respiration belt, PPG sensors). *2) Absolute Error (AE).* We quantify the breathing and heartbeat rate estimation performance using absolute error (AE), which is defined as $AE = |R_m - R_g|$, where R_m and R_g denote the breathing/heartbeat rate or body fat ratio measured by our framework and the ground truth from medical sensors, respectively. Specifically, we utilize beats per minute (BPM) to measure the breathing/heartbeat rate in our evaluations and compute number of breathing/heartbeat cycles missed in each measurement as corresponding absolute errors, which intuitively reflects the difference between our breathing and heartbeat rate measurement and ground truth rates captured by medical sensors. *3) Accuracy.* For user identification and gender recognition modules, the identification/recognition accuracy is used to evaluate the user identification and gender recognition performance, which denotes the percentage of data segments correctly identified or recognized as belonging to the correct labels (e.g., user class and gender label). *4) R^2 score.* We evaluate the performance of body fat percentage estimation using R^2 score, which is used to determine whether the correlation between the ground truth and estimated fat percentage exists. We define $R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$, where y refers to the ground truth of body fat percentage and f represents the regression results. R^2 score ranges from 0 to 1 and the correlation exists if R^2 score is higher than 0.5.

7.2 Performance of Breathing and Heartbeat Waveform Reconstruction

To evaluate the performance of vital sign signal reconstruction, we measure RMSEs between the reconstructed waveforms (RW) and the ground truth. We present same-user performance (the dataset is shuffled and split for training and testing) performance and cross-user performance (80% users for training and other 20% users for testing). The results are shown in Table 1. We compare the RMSEs with bench facial vibrations passing bandpass filters with cut-off frequencies matching breathing and heartbeat frequencies.

Performance on Standalone Headsets. For Meta Quest, the average RMSEs of breathing/heartbeat pattern reconstruction are 0.14/0.46 and 1.73/3.50 for same-user and cross-user scenarios. Reconstruction examples are shown in Figure 13. We observe that our framework can successfully reconstruct breathing and heartbeat waveform from pre-selected readings, which exhibits similar



(a) Facial vibration (0.1~0.5Hz) (b) Facial vibration (0.8~3.0Hz)
(c) Reconstructed signal (overall) (d) Reconstructed signal (overall)
(e) Reconstructed signal (cross-user) (f) Reconstructed signal (cross-user)

Figure 13: Examples of vital sign signal reconstruction (x-axis, y-axis of gyroscope and z-axis of accelerometer) and heartbeat (x-axis of gyroscope and y-axis, z-axis of accelerometer) using Meta Quest.

patterns compared to the ground truths. For HTC Vive Pro Eye, the RMSEs achieve 0.32/0.38 and 2.10/4.95, which show that our framework is effective even under challenging cross-user settings.

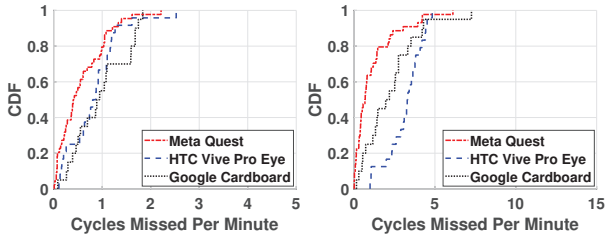
Performance on Low-end Headsets. For low-cost Google Cardboard headsets, the RMSEs of breathing/heartbeat derivation are all below 0.31/0.79 and 2.61/5.72. It demonstrates that our framework still has good performance with a lower sampling rate and less sensitive motion sensors of smartphones. Overall, after waveform reconstruction, RMSEs improve dramatically, which illustrates the effectiveness of our design on breathing/heartbeat waveform reconstruction.

7.3 Performance of Breathing and Heartbeat Rate Estimation

To evaluate the performance of breathing and heartbeat rate estimation, we calculate the AEs between our breathing/heartbeat measurement and ground truth rate captured by medical devices in the four aforementioned scenarios. The results are shown in Figure 14 leveraging cumulative distribution functions (CDF).

Performance on Standalone Headsets. Regarding breathing rate estimation with Meta Quest, we can observe from Figure 14(a) that the AEs are less than 2.0*BPM*, which indicates that less than 2 cycles of human's respiration are missed with our proposed breathing and heartbeat rate estimation module. For the results heartbeat rate estimation on Meta Quest in Figure 14(b), the AEs are usually less than 5.0*BPM*. As for the performance of breathing and heartbeat rate estimation using HTC Vive Pro Eye, we also achieve comparable performance, with less than 2.5*BPM* and 5.0*BPM* AEs for most measurements. The results demonstrate that our framework can achieve accurate of estimation of vital sign rates under various scenarios leveraging standalone AR/VR headsets.

Performance on Low-end Headsets. From the results displayed from Figure 14(a) and Figure 14(b), while the performance of vital sign rate estimation using Google Cardboard with Samsung Galaxy S6 has some degradation compared to standalone AR/VR headsets, the AEs of breathing and heartbeat rate estimation can also reach less than 2.0*BPM* and 7.5*BPM* is most cases of



(a) Absolute error of breathing (b) Absolute error of heartbeat

Figure 14: Cumulative Distribution Functions (CDF) on Absolute Errors (AEs) of Beats Per Minute (BPM) from breathing and heartbeat rate estimation.

measurements. The results manifest that our proposed breathing and heartbeat rate estimation module can also perform well upon low-end AR/VR headsets.

7.4 Performance of User Identification

For user identification, facial vibrations induced by breathing and heartbeat are utilized separately to evaluate the performance. The ratio of training and testing size is set to be 8 : 2. The performance of user identification is shown in Figure 15(a).

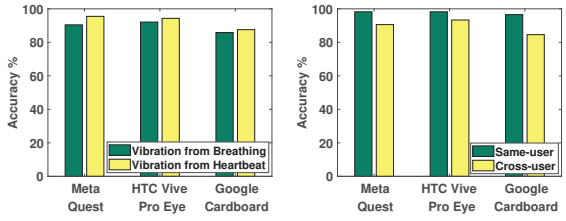
Performance on Standalone Headsets. For Meta Quest, our framework has achieved 90.43% accuracy using vibrations from breathing, and 95.51% accuracy with vibrations from heartbeat. Our framework also achieve more than 92.11% and 94.32% on user identification on HTC Vive. High identification accuracy on two types of commodity standalone AR/VR headsets demonstrates the effectiveness of our framework on differentiate users.

Performance on Low-end Headsets. For Google Cardboard, our framework can achieve more than 85.83% and 87.56% accuracy leveraging breathing and heartbeat vibrations, respectively. Compared to standalone AR/VR headsets, we find that cardboard devices performs slightly worse for user identification. The reason could be the hard cardboard device cannot well fit all face shapes of different users, rendering less intensive facial vibrations induced by vital signs. Nevertheless, the results have exhibited potentials of realizing user identification with less sensitive sensors on off-the-shelf low-end headsets.

7.5 Performance of Gender Detection

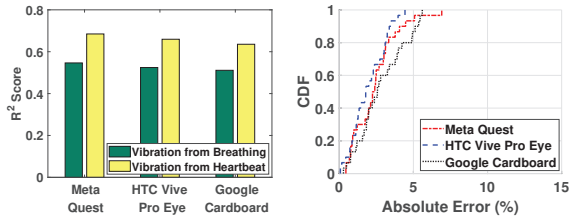
We evaluate the gender detection performance for the three AR/VR headsets. Figure 15(b) shows the gender recognition accuracy of both same-user scenario (the user exists in the training data) and cross-user testing (testing on new users).

Performance on Standalone Headsets. For Meta Quest, the framework has achieved 98.23% accuracy for gender recognition tasks on same-user setting and 90.57% accuracy in cross-user setting. For HTC Vive Pro Eye, the framework can achieve more than 98.23% and 93.33% accuracy in both settings, respectively. The results demonstrate the robustness of gender-related hidden feature extraction, even under the realistic and challenging cross-user setting, and further indicate the effectiveness of our proposed framework on gender detection.



(a) User identification accuracy (b) Gender detection accuracy

Figure 15: Recognition accuracy of user identification and gender detection.



(a) R^2 score of cross-validation (b) Error of cross-user testing

Figure 16: R^2 score and error rate of cross-user evaluation on body fat percentage estimation.

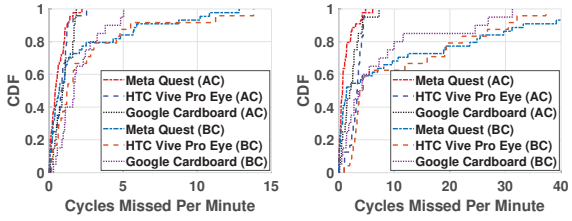
Performance on Standalone Headsets. For Google Cardboard, the framework can still achieve more than 96.54% and 84.58% accuracy on same-user and cross-user scenarios. The high gender detection accuracy shows that our system will also work well with smartphone-based AR/VR headsets, which only have low-grade motion sensors with limited sampling rates.

7.6 Performance of Body Fat Percentage Estimation

To evaluate the performance of body fat percentage estimation, R^2 scores are computed with cross-validation lasso regression. We also evaluate the generalization performance by testing the regression model on cross-user scenario. The results are shown in Figure 16.

Performance on Standalone Headsets. Using facial vibrations induced by human's breathing, the cross-validation R^2 scores are 0.5464 and 0.5245, 0.5112 for Meta Quest and HTC Vive Pro Eye. Utilizing heartbeat-induced vibrations, R^2 scores can reach more than 0.6848 and 0.6595 for two commodity standalone AR/VR headsets. We also evaluate the performance of the body fat percentage estimation module on a group of samples from new users and the results are shown in Figure 16. From the evaluation results of Meta Quest and HTC Vive Pro Eye, the absolute error of body fat ratio estimation is less than 2.6% in most measurements. The results manifest that there exist correlations between fat percentage and facial vibrations features and our proposed method can perform well on estimating body fat percentage.

Performance On Low-end Headsets. For Google Cardboard, the cross-validation R^2 scores are 0.5112 and 0.6536 for breathing- and heartbeat-induced vibrations. The absolute error for cross-user testing is less than 5.0% in most cases, which illustrates the possibility of estimating body fat percentage using low-end AR/VR headsets.



(a) Absolute error of breathing (b) Absolute error of heartbeat

Figure 17: Absolute errors of breathing and heartbeat estimation before (BC) and after motion artifact cancellation (AC).

7.7 Evaluation on Framework Robustness

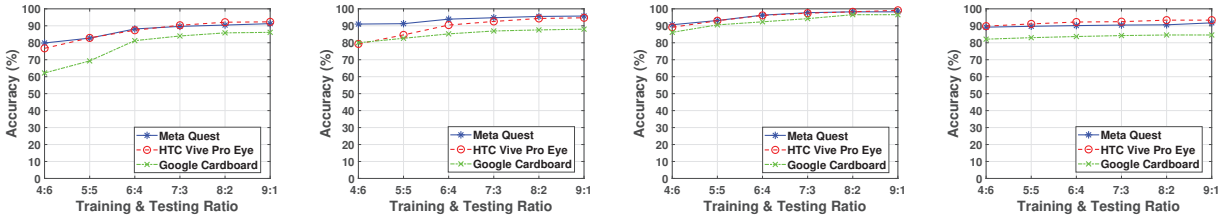
To evaluate the robustness of the framework, we consider two impacting factors and test its performance, including effects of motion artifact cancellation and training size.

Effect of Motion Artifact Cancellation. To study the impact of our proposed motion artifact cancellation scheme, we compute the absolute errors of breathing and heartbeat rate estimation before and after motion artifact cancellation. As shown in Figure 17, we can observe some measurements with large absolute errors for breathing and heartbeat rate estimation. After cancellation, the AEs of breathing and heartbeat estimation have significant drops for all 3 types of commodity AR/VR headsets, which demonstrate the effectiveness of our motion artifact cancellation scheme.

Effect of Training Size. Intuitively, if the framework can achieve high accuracy with less training samples, it demonstrates that users can register for short time to function the framework. Thus, we investigate the influence of training samples to demonstrate the efficiency of our proposed framework. Specifically, we evaluate user identification and gender recognition red with different training samples, which is shown in Figure 18. It can be observed that the framework still retains more than 88.13% and 90.11% accuracy for user identification and gender detection with the training size of 6:4, which demonstrates that the framework will still work well despite of short-time data collection from users.

8 RELATED WORKS

Vital Sign and Health Status Monitoring. Approaches with different sensors have been explored for acquiring vital sign information. Earlier works [55] showed the use of PPG and ECG sensors to for heart rate monitoring. In addition, recent studies have demonstrated the use of Radio Frequency (RF) for capturing vital sign information. Some of the representative studies explore the use of off-the-shelf WiFi infrastructures to measure vital signs. For example, Liu et al. [28] propose to track the breathing rate and heartbeat of users during sleep. As another example, PhaseBeat [63] exploits the phase information in WiFi signals to sense heart beating. Zheng et al. [67] present V^{2}iFi that performs vital sign monitoring using impulse radio signals. MoVi-Fi [] innovatively employs approaches based on deep contrastive learning to recover fine-grained vital signs waveform under major body movements, and is demonstrated to perform well under various and complicated practical scenarios. Although existing works have explored the potential of health status monitoring through different ways,



(a) Accuracy from breathing (b) Accuracy from heartbeat (c) Same-user testing (d) Cross-user testing

Figure 18: User identification and gender recognition accuracy with different numbers of training samples.

our work is the first health status monitoring framework deployed on commodity AR/VR headsets leveraging built-in motion sensors.

Authentication on AR/VR Headsets. With the drastic development of AR/VR applications, AR/VR authentications have also experienced a presperity of rapid development and generation. The work by Yada et al. [66] using PIN authentication in AR systems accomplished encouraging results. A recent study by George et al. [22] evaluated the security and usability of SWIPE and PIN within the VR environment and found that usability was comparable in performance to the mobile version of PIN and SWIPE. Furthermore, LookUnlock [21] allows users to input passwords by looking at virtual objects in an encoded sequence. Wazir et al. [65] demonstrated that it was also possible to recognize users using AR based on their writing patterns. In addition, NodtoAuth [64], a AR/VR authentication system, was proposed to allow users to unlock their headsets by nodding. Shi et al. [52] demonstrate a vulnerability where user identification can be accomplished using facial dynamics captured by the motion sensors inside AR/VR headsets. In our work, personalization can be achieved without specific interactions with AR/VR devices, which is more convenient in daily usage.

9 DISCUSSION AND FUTURE WORKS

While our proposed vital sign monitoring system leveraging AR/VR headsets showcases its potential in various healthcare functions, such as breathing and heartbeat waveform reconstruction, breathing and heartbeat rate estimation, user identification, gender recognition and body fat ratio estimation, there is still much room for improvement in terms of effectiveness, accuracy and robustness.

Involvement of Continuous Motions. In this work, we design an adaptive filter to dynamically cancel the impacts of motion artifacts in facial vibrations. As the designed filter requires using segments of reference signals (i.e., relatively static scenarios), it may not work well under continuous motions in some rare AR/VR scenarios, such as continuously walking and running. However, these scenarios are not common since AR/VR systems normally constrain the size of the play area, such as the Guardian in Meta Quest. The user is only allowed to operate the headset and the controllers within such areas. Moreover, it is always possible for our proposed system to localize a signal segment of relatively static posture before the continuous motions (e.g, a short-time break during a VR game) as the reference signal. With such considerations, our framework can still capture appropriate reference signals for motion artifact cancellation. In the future, we plan to explore some advanced deep-learning-based body movement artifact removal techniques to improve the framework robustness under such extreme cases with long-lasting and continuous motions. We will also consider learning-based techniques to mitigate smaller-scale face motions, such as talking, chewing, and laughing.

Improvements on Body Fat Ratio Estimation. Regarding the body fat ratio estimation module, our current study has involved 10 users with body fat ratio ranging from 9% to 28%, and monitors their body fat for 1 month, which has provided us with initial insights into the potential of our system on body fat ratio estimation. To further validate and generalize our findings, we plan to include a more diverse group of participants and monitor their body fat changes over longer periods of time. We also plan to involve human subjects with a wider range of body fat ratios and employ commodity medical devices (e.g., InBody 570 Body Composition Analyzer [19]) to acquire the ground truth, thus to improve the robustness of our body fat percentage estimation module. This will allow us to better understand the system's performance across a wider range of body types and further enhance the accuracy of our body fat ratio estimation design.

10 CONCLUSION

In this paper, we propose a health monitoring framework on commodity AR/VR headsets, which utilizes vital-sign-induced facial vibrations captured by motion sensors. We design an adaptive-filter-based method to remove motion artifacts and propose a vibration enhancement strategy via axis fusion. To realize healthcare functions, we reconstruct precise breathing/heartbeat waveforms through a LSTM-based model with self-attention scheme, with distinctive biomarker extraction. We show the scalability of our framework on four representative applications, including breathing/heartbeat rate estimation, user identification, gender detection, and body fat percentage estimation. We validate our framework via extensive experiments on 3 different types of commodity AR/VR headsets and demonstrate its effectiveness, robustness and generality. We believe that our proposed framework will deliver innovative ideas of smart and personalized healthcare functions in AR/VR community, thus facilitates the development of more practical healthcare services on commodity AR/VR devices.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation Grants CCF2211163, CCF2120396, CNS2120276, CCF2000480, CNS2145389, OAC2139358, CNS2201465 and CNS2152669.

REFERENCES

[1] 2019. Vital Signs (Body Temperature, Pulse Rate, Respiration Rate, Blood Pressure). https://www.hopkinsmedicine.org/health/conditions-and-diseases/vital-signs-body-temperature-pulse-rate-respiration-rate-blood-pressure.

[2] 2020. What's a normal resting heart rate? - Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/.

[3] 2021. Gender and health. https://www.who.int/news-room/questions-and-answers/item/gender-and-health.

[4] 2022. Metaverse. https://en.wikipedia.org/wiki/Metaverse.

[5] 2022. Virtual Reality Market Size Worth USD 43.01 Billion in 2028. https://www.emergenresearch.com/press-release/global-virtual-reality-market.

[6] 2022. Virtual reality sickness. https://en.wikipedia.org/wiki/Virtual_reality_sickness.

[7] M. Akman, M.K. Uçar, Z. Uçar, K. Uçar, B. Baraklı, and M.R. Bozkurt. 2022. Determination of Body Fat Percentage by Gender Based with Photoplethysmography Signal Using Machine Learning Algorithm. *IRBM* 43, 3 (2022), 169–186. https://doi.org/10.1016/j.irbm.2020.12.003

[8] Musaed Alhussein and Ghulam Muhammad. 2018. Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework. *IEEE Access* 6 (2018), 41034–41041. https://doi.org/10.1109/ACCESS.2018.2856238

[9] S Abhishek Anand and Nitesh Saxena. 2018. Speechless: Analyzing the Threat to Speech Privacy from Smartphone Motion Sensors. In *2018 IEEE Symposium on Security and Privacy (SP)*. 1000–1017. https://doi.org/10.1109/SP.2018.00004

[10] Anatomy and Physiology. 2019. Cardiac Cycle. http://library.open.oregonstate.edu/aandp/chapter/19-3-cardiac-cycle/.

[11] Jeffrey Bloom, Michael J. Lopez, and Appaji Rayi. 2021. *Anatomy, Head and Neck, Eye Levator Labii Superioris Muscle*. StatPearls Publishing, Treasure Island (FL). http://europepmc.org/books/NBK541031

[12] Google Cardboard. 2014. Cardboard - Google VR. https://arvr.google.com/cardboard/.

[13] Eunhee Chang, Hyun Taek Kim, and Byounghyun Yoo. 2022. Identifying physiological correlates of cybersickness using heartbeat-evoked potential analysis. *Virtual Reality* 26, 3 (01 Sep 2022), 1193–1205. https://doi.org/10.1007/s10055-021-00622-2

[14] ClassVR. 2022. ClassVR: virtual reality for schools. https://www.classvr.com/us/.

[15] Hao Dang, Muyi Sun, Guanhong Zhang, Xingqun Qi, Xiaoguang Zhou, and Qing Chang. 2019. A Novel Deep Arrhythmia-Diagnosis Network for Atrial Fibrillation Classification Using Electrocardiogram Signals. *IEEE Access* 7 (2019), 75577–75590. https://doi.org/10.1109/ACCESS.2019.2918792

[16] Anne E Dixon and Ubong Peters. 2018. The effect of obesity on lung function. *Expert Rev Respir Med* 12, 9 (Aug. 2018), 755–767.

[17] VR Doctors. 2022. VR Doctors – change reality, change healthcare. http://vrdoctors.net/.

[18] Biyi Fang, Nicholas D. Lane, Mi Zhang, Aidan Boran, and Fahim Kawsar. 2016. BodyScan: Enabling Radio-Based Sensing on Wearable Devices for Contactless Activity and Vital Sign Monitoring *(MobiSys '16)*. Association for Computing Machinery, New York, NY, USA, 97–110. https://doi.org/10.1145/2906388.2906411

[19] Recovery for Athletes. 2022. InBody 570 Body Composition Analyzer. https://www.recoveryforathletes.com/products/inbody-570-body-composition-analyzer.

[20] Jann Philipp Freiwald, Yvonne Göbel, Fariba Mostajeran, and Frank Steinicke. 2020. The Cybersickness Susceptibility Questionnaire: Predicting Virtual Reality Tolerance. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 115–118. https://doi.org/10.1145/3404983.3410022

[21] Markus Funk, Karola Marky, Iori Mizutani, Mareike Kritzler, Simon Mayer, and Florian Michahelles. 2019. Lookunlock: Using spatial-targets for user-authentication on hmds. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[22] Ceenu George, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. NDSS.

[23] Azadeh Hadadi, Christophe Guillet, Jean-Rémy Chardonnet, Mikhail Langovoy, Yuyang Wang, and Jivka Ovtcharova. 2022. Prediction of cybersickness in virtual environments using topological data analysis and machine learning. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.973236

[24] HealthData.gov. 2022. HealthData.gov. https://healthdata.gov//.

[25] Andrii Horiachko. 2022. AR and VR for Video Conferencing. https://www.softermii.com/blog/ar-and-vr-for-video-conferencing.

[26] Lennart Leicht, Pascal Vetter, Steffen Leonhardt, and Daniel Teichmann. 2017. The PhysioBelt: A safety belt integrated sensor system for heart activity and respiration. In *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. 191–195. https://doi.org/10.1109/ICVES.2017.7991924

[27] Boning Li and Akane Sano. 2020. Extraction and Interpretation of Deep Autoencoder-Based Temporal Features from Wearables for Forecasting Personalized Mood, Health, and Stress. 4, 2, Article 49 (jun 2020), 26 pages. https://doi.org/10.1145/3397318

[28] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking Vital Signs During Sleep Leveraging Off-the-Shelf WiFi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (Hangzhou, China) *(MobiHoc '15)*. Association for Computing Machinery, New York, NY, USA, 267–276. https://doi.org/10.1145/2746285.2746303

[29] Martin László, Zoltán Istenes, and Adam Tarcsi. 2020. Extracting Physiological Signals from Smartphone Sensors.

[30] Francisco Martín-Rodríguez, José Luis Martín Conty, Verónica Casado Vicente, Pedro Arnillas Gómez, Alicia Mohedano-Moriano, and Miguel Ángel Castro Villamor. 2018. Does gender influence physiological tolerance in resuscitators when using personal protection equipment against biological hazards? *Emergency Medicine International* 2018 (2018), 1–7. https://doi.org/10.1155/2018/5890535

[31] Carey R. Merritt, H. Troy Nagle, and Edward Grant. 2009. Textile-Based Capacitive Sensors for Respiration Monitoring. *IEEE Sensors Journal* 9, 1 (2009), 71–78. https://doi.org/10.1109/JSEN.2008.2010356

[32] Meta. 2021. How AR is making video calling more collaborative. https://tech.facebook.com/reality-labs/2021/12/how-ar-is-making-video-calling-more-collaborative/.

[33] Meta. 2021. Unleash your creativity through VR art, paintings sculptures. https://www.oculus.com/blog/unleash-your-creativity-through-vr-art//.

[34] Abdul Momin, Saheli Bhattacharya, Sudip Sanyal, and Pavan Chakraborty. 2020. Visual Attention, Mental Stress and Gender: A Study Using Physiological Signals. *IEEE Access* 8 (2020), 165973–165988. https://doi.org/10.1109/ACCESS.2020.3022727

[35] Neulog. 2011. Heart Rate Pulse logger sensor NUL-208. https://neulog.com/heart-rate-pulse/.

[36] Neulog. 2011. Respiration Monitor Belt logger sensor NUL-236. https://neulog.com/respiration-monitor-belt/.

[37] Keith Nolan, Aidan Mooney, and Susan Bergin. 2019. An Investigation of Gender Differences in Computer Science Using Physiological, Psychological and Behavioural Metrics. In *Proceedings of the Twenty-First Australasian Computing Education Conference* (Sydney, NSW, Australia) *(ACE '19)*. Association for Computing Machinery, New York, NY, USA, 47–55. https://doi.org/10.1145/3286960.3286966

[38] Oculus. 2020. Oculus Quest Store: VR Games, Apps, and More. https://www.oculus.com/experiences/quest/.

[39] Heeseok Oh and Wookho Son. 2022. Cybersickness and Its Severity Arising from Virtual Reality Content: A Comprehensive Study. *Sensors (Basel)* 22, 4 (Feb. 2022).

[40] Donald B. Percival and Andrew T. Walden. 1993. *Spectral Analysis for Physical Applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511622762

[41] Richard A. Davis Peter J. Brockwell. 1991. *Time Series: Theory and Methods*. Springer New York, NY. https://doi.org/10.1007/978-1-4419-0320-4

[42] Caitlin Polley, Titus Jayarathna, Upul Gunawardana, Ganesh Naik, Tara Hamilton, Emilio Andreozzi, Paolo Bifulco, Daniele Esposito, Jessica Centracchio, and Gaetano Gargiulo. 2021. Wearable Bluetooth Triage Healthcare Monitoring System. *Sensors* 21, 22 (2021). https://doi.org/10.3390/s21227586

[43] Possible. 2022. Can Excess Weight Have Major Impact On Your Blood Circulation? http://https://possible.in/can-excess-weight-have-major-impact-on-your-blood-circulation.html/.

[44] Chenxin Qu, Xiaoping Che, Siqi Ma, and Shuqin Zhu. 2022. Bio-physiological-signals-based VR cybersickness detection. *CCF Transactions on Pervasive Computing and Interaction* 4, 3 (01 Sep 2022), 268–284. https://doi.org/10.1007/s42486-022-00103-8

[45] RENPHO. 2022. Best Body Fat and Body Composition Analyzer Scale, Renpho. https://renpho.com/collections/renpho-scales.

[46] S. Sabeeha and C. Shiny. 2017. ECG-based heartbeat classification for disease diagnosis. In *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. 1113–1117. https://doi.org/10.1109/ICCMC.2017.8282646

[47] E. Sardini and M. Serpelloni. 2010. Instrumented wearable belt for wireless health monitoring. *Procedia Engineering* 5 (2010), 580–583. https://doi.org/10.1016/j.proeng.2010.09.176 Eurosensor XXIV Conference.

[48] Ali H. Sayed. 2003. Fundamentals Of Adaptive Filtering.

[49] M. Scarpetta, M. Spadavecchia, G. Andria, M.A. Ragolia, and N. Giaquinto. 2022. Accurate simultaneous measurement of heartbeat and respiratory intervals using a smartphone. 17, 07 (jul 2022), P07020. https://doi.org/10.1088/1748-0221/17/07/P07020

[50] Volkan Sevinc and Mehmet Ilker Berkman. 2020. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Applied Ergonomics* 82 (2020), 102958. https://doi.org/10.1016/j.apergo.2019.102958

[51] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-Attention with Relative Position Representations. https://doi.org/10.48550/ARXIV.1803.02155

[52] Cong Shi, Xiangyu Xu, Tianfang Zhang, Payton Walker, Yi Wu, Jian Liu, Nitesh Saxena, Yingying Chen, and Jiadi Yu. 2021. Face-Mic: Inferring Live Speech and Speaker Identity via Subtle Facial Dynamics Captured by AR/VR Motion

Sensors *(MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 478–490. https://doi.org/10.1145/3447993.3483272

[53] Abdulhamit Subasi, Mariam Radhwan, Rabea Kurdi, and Kholoud Khateeb. 2018. IoT based mobile healthcare system for human activity recognition. In *2018 15th Learning and Technology Conference (LT)*. 29–34. https://doi.org/10.1109/LT.2018.8368507

[54] Tim J.T. Sutherland, Christene R. McLachlan, Malcolm R. Sears, Richie Poulton, and Robert J. Hancox. 2016. The relationship between body fat and respiratory function in young adults. *European Respiratory Journal* 48, 3 (2016), 734–747. https://doi.org/10.1183/13993003.02216-2015 arXiv:https://erj.ersjournals.com/content/48/3/734.full.pdf

[55] H. Emrah Tasli, Amogh Gudi, and Marten den Uyl. 2014. Remote PPG based vital sign measurement using adaptive facial regions. In *2014 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/ICIP.2014.7025282

[56] Osamu Tochikubo, Eiji Miyajima, Tomohiko Shigemasa, and Masao Ishii. 1999. Relation Between Body Fat–Corrected ECG Voltage and Ambulatory Blood Pressure in Patients With Essential Hypertension. *Hypertension* 33, 5 (May 1999), 1159–1163. https://doi.org/10.1161/01.HYP.33.5.1159

[57] UltraLeap. 2022. What is VR training? https://www.ultraleap.com/company/news/blog/vr-training/.

[58] Muhammed Kürşad Uçar, Zeliha Uçar, Kübra Uçar, Mehmet Akman, and Mehmet Recep Bozkurt. 2021. Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence. *Biomedical Signal Processing and Control* 68 (2021), 102650. https://doi.org/10.1016/j.bspc.2021.102650

[59] Michael Vallance and Phillip A. Towndrow. 2022. Perspective: Narrative Storyliving in Virtual Reality Design. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.779148

[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv:1706.03762 [cs.CL]

[61] HTC Vive. 2019. VIVE Pro Eye Overview, VIVE United States. https://www.vive.com/us/product/vive-pro-eye/overview/.

[62] Danling Wang, Qifeng Zhang, Md. Razuan Hossain, and Michael Johnson. 2018. High Sensitive Breath Sensor Based on Nanostructured K2W7O22 for Detection of Type 1 Diabetes. *IEEE Sensors Journal* 18, 11 (2018), 4399–4404. https://doi.org/10.1109/JSEN.2018.2825302

[63] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI Phase Data for Vital Sign Monitoring with Commodity WiFi Devices. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. 1230–1239. https://doi.org/10.1109/ICDCS.2017.206

[64] Xue Wang and Yang Zhang. 2021. Nod to Auth: Fluent AR/VR Authentication with User Head-Neck Modeling. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[65] Waqas Wazir, Hasan Ali Khattak, Ahmad Almogren, Mudassar Ali Khan, and Ikram Ud Din. 2020. Doodle-based authentication technique using augmented reality. *IEEE Access* 8 (2020), 4022–4034.

[66] Dhruv Kumar Yadav, Beatrice Ionascu, Sai Vamsi Krishna Ongole, Aditi Roy, and Nasir Memon. 2015. Design and analysis of shoulder surfing resistant pin based authentication mechanisms on google glass. In *International conference on financial cryptography and data security*. Springer, 281–297.

[67] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. 2020. V2iFi: In-Vehicle Vital Sign Monitoring via Compact RF Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 70 (jun 2020), 27 pages. https://doi.org/10.1145/3397321

,