FaceReader: Unobtrusively Mining Vital Signs and Vital Sign Embedded Sensitive Info via AR/VR Motion Sensors

Tianfang Zhang
Rutgers University
tz203@scarletmail.rutgers.edu

Zhengkun Ye
Temple University
zhengkun.ye@temple.edu

Ahmed Tanvir Mahdad
Texas A&M University
mahdad@tamu.edu

Md Mojibur Rahman Redoy Akanda
Texas A&M University
redoy.akanda@tamu.edu

Cong Shi
New Jersey Institute of Technology
cong.shi@njit.edu

Yan Wang
Temple University
y.wang@temple.edu

Nitesh Saxena
Texas A&M University
nsaxena@tamu.edu

Yingying Chen*
Rutgers University
yingche@scarletmail.rutgers.edu

ABSTRACT

The market size of augmented reality and virtual reality (AR/VR) has been expanding rapidly in recent years, with the use of face-mounted headsets extending beyond gaming to various application sectors, such as education, healthcare, and the military. Despite the rapid growth, the understanding of information leakage through sensor-rich headsets remains in its infancy. Some of the headset's built-in sensors do not require users' permission to access, and any apps and websites can acquire their readings. While these *unrestricted* sensors are generally considered free of privacy risks, we find that an adversary could uncover private information by scrutinizing sensor readings, making existing AR/VR apps and websites potential eavesdroppers. In this work, we investigate a novel, unobtrusive privacy attack called *FaceReader*, which reconstructs high-quality vital sign signals (breathing and heartbeat patterns) based on unrestricted AR/VR motion sensors. *FaceReader* is built on the key insight that the headset is closely mounted on the user's face, allowing the motion sensors to detect subtle facial vibrations produced by users' breathing and heartbeats. Based on the reconstructed vital signs, we further investigate three more advanced attacks, including gender recognition, user re-identification, and body fat ratio estimation. Such attacks pose severe privacy concerns, as an adversary may obtain users' sensitive demographic/physiological traits and potentially uncover their real-world identities. Compared to prior privacy attacks relying on speeches and activities, *FaceReader* targets spontaneous breathing and heartbeat activities that are naturally produced by the human body and are unobtrusive to victims. In particular, we design an adaptive filter to dynamically mitigate the impacts of body motions. We further employ advanced deep-learning techniques to reconstruct vital sign signals, achieving signal qualities comparable to those of dedicated medical instruments, as well as deriving sensitive gender, identity, and body fat information. We conduct extensive experiments involving 35 users on three types of mainstream AR/VR headsets across 3 months. The results reveal that *FaceReader* can reconstruct vital signs with low mean errors and accurately detect gender (over 93.33%). The attack can also link/re-identify users across different apps, websites, and longitudinal sessions with over 97.83% accuracy. Furthermore, we present the first successful attempt at revealing body fat information from motion sensor data, achieving a remarkably low estimation error of 4.43%.

CCS CONCEPTS

• **Security and privacy** → ***Hardware attacks and countermeasures***.

KEYWORDS

Vital sign, Sensitive info, AR/VR headsets, Motion sensors

ACM Reference Format:
Tianfang Zhang, Zhengkun Ye, Ahmed Tanvir Mahdad, Md Mojibur Rahman Redoy Akanda, Cong Shi, Yan Wang, Nitesh Saxena, and Yingying Chen. 2023. FaceReader: Unobtrusively Mining Vital Signs and Vital Sign Embedded Sensitive Info via AR/VR Motion Sensors. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23), November 26–30, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3576915.3623102

1 INTRODUCTION

Augmented reality/virtual reality (AR/VR) technologies have created an emerging market: the hardware and software revenues top 31.12 billion in 2023 and have a projected market volume of 52.05 billion in 2027 [39]. AR/VR face-mounted headsets have grown in popularity within the gaming community, and their use has been extended to education [7], healthcare [44], military [40], and beyond. Despite the transformative potential of AR/VR, the developments of corresponding privacy protection techniques and policies remain in their infancy. Recent studies highlight privacy leakages through active virtual engagements [13, 27], where adversaries extract private information through users' direct text input in AR/VR scenarios (e.g., containing users' names, gender, and shopping preferences). These attacks typically employ cognitive tricks during virtual conversations, such as social engineering tactics, self-disclosure mechanisms, and adversarial AR/VR game designs, to deceive users into disclosing their private information.

*Yingying Chen is the corresponding author.

CCS '23, November 26–30, 2023, Copenhagen, Denmark

ACM ISBN 979-8-4007-0050-7/23/11…$15.00
https://doi.org/10.1145/3576915.3623102

However, we find that a variety of unrestricted sensors in AR/VR headsets can enable a more stealthy form of privacy attack that does not involve any active engagement from the user. The unrestricted sensors (e.g., accelerometer, gyroscope), which are designed to utilize users' body movements to support immersive human-computer interactions, are always on by default. This configuration provides opportunities for passive attacks, where adversaries can infer private attributes from sensor data during normal AR/VR activities (e.g., gaming, surfing websites), without requiring any direct interactions or cognitive tricks. Such private information could be utilized for various malicious purposes, such as cyber-tracking [35], targeted/personalized advertising [35], curtailing personal autonomy [6], cyber abuse [11], or even advancing political agendas [22]. The attacks using unrestricted sensors can significantly reduce the risk of exposing the malicious intent, thereby achieving a higher degree of stealthiness compared to active virtual engagements. In this work, we explore such passive attacks by sensing facial vibrations induced by vital signs of users (i.e., breathing and heart beating) through unrestricted AR/VR motion sensors. The designed attack can lead to severe privacy leakage, revealing details such as gender, identity, lifestyle, preferences, and even body fat ratio, without assuming any active inputs from the users.

Proposed *FaceReader* Based on Vital Sign Reconstruction. We find that human breathing and heartbeats, which correlate with contractions and relaxations of the nasal cavity and facial blood vessels, produce *minute facial vibrations* on the face surface. Since AR/VR headsets are tightly mounted on users' heads and touch various facial areas (e.g., forehead, nose, temples), these vibrations can propagate through the headset and vibrate the motion sensors. Based on this insight, we design a new passive privacy attack, *FaceReader*, which reconstructs high-quality vital signs, specifically breathing and heartbeat signals, from motion sensor readings. These fine-grained vital signs contain rich biometrics and thus can be leveraged to realize more advanced attacks, deriving users' private attributes such as gender, identity, and body fat ratio. As *FaceReader* only requires accessing unrestricted motion sensors, the consequences could be dire if such privacy attacks are launched on a large scale by enterprises or malicious actors to acquire comprehensive profiles of a large number of AR/VR users. We demonstrate that two levels of privacy leakages are possible via *FaceReader*:

Direct attack: vital sign reconstruction. We demonstrate the feasibility of reconstructing vital sign signals that yield similar signal quality to medical instruments, such as Photoplethysmogram (PPG) sensors and respiration monitoring belts (i.e., NeuLog NUL-236 respiration monitor belt). The reconstructed patterns of breathing and heartbeats can be linked to important heart and lung functions, such as the metabolism of oxygen, glucose, and lung capacity. Thus, the fine-grained vital sign information can be categorized as highly private information by USA GOV [15], and disclosure of such information could lead to medical discrimination and fraud of healthcare records (more consequences are elaborated in Section 2.1).

Advanced attacks: sensitive information derivation. Based on the high-quality vital signs, we show more profound attacks that reveal sensitive information embedded within or extracted from vital sign signals: *(1) Gender Recognition:* Our attack could reveal users' gender information, which could be exploited to push gender-specific advertisements [35] or perform romance scams. *(2) User Re-identification:* With the vital sign signals containing rich biometric information (e.g., related to functions of the heart and lungs), adversaries potentially link/re-identify users across apps, websites, and longitudinal sessions. The attack is particularly concerning as more users frequently post sensitive data (e.g., game reviews and photos) on multiple websites, and the identity linkage could expose their preferences, lifestyles, and even real-world identities.

(3) Body Fat Ratio Estimation: The reconstructed vital signs can be leveraged to estimate the body fat ratio, which affects respiratory resistance and blood flow of users [4, 8]. Adversaries can leverage such information to effectively manipulate users' minds (e.g., promoting health-related products [3], performing cyber-bullying [30]) or sell the information for profit.

Challenges of Deriving Privacy Information via Facial Vibrations. To successfully execute our proposed attack leveraging built-in motion sensors on commodity AR/VR headsets, we face several practical challenges: *(1) Capturing Subtle Facial Vibrations on AR/VR Headsets:* The amplitude of these vibrations induced by human respiration and heartbeat are very small, while motion sensors are designed to capture large-scale body motions, making it difficult to sense fine-grained vital signs. *(2) Interference of Motion Artifact:* In AR/VR scenarios, users interact with the virtual objects and environments with hand, head, and body motions, which cause significant distortions to the minute vital sign patterns. It is challenging to separate vital-sign-induced facial vibrations from the motion artifact. Therefore, we need to design effective techniques to separate facial vibrations from other types of body movements. *(3) Extract Effective Features for Advanced Attacks:* Vital sign signals are encoded wide range of biometrics. To realize effective advanced attacks, we need to design features that capture task-specific characteristics while suppressing the others.

Our Technical Contributions. Taking motion sensor readings as input, our attack first mitigates motion artifacts using an adaptive filtering algorithm. This algorithm maintains signals exhibiting strong periodicity and resembling breathing and heartbeat cycles, while simultaneously removing other irregular motion patterns. In addition, we design a recurrent neural network with attention mechanisms to reconstruct fine-grained vital sign signals from the time-frequency features of facial vibrations, achieving signal quality comparable to dedicated medical instruments. Given the reconstructed signals, we highlight the potential privacy risks by developing features and recognition methods for body fat ratio estimation, user re-identification, and gender recognition. Our contributions are summarized as follows:

- We discover a new passive privacy attack that reconstructs high-quality vital sign signals from unrestricted AR/VR motion sensors. The proposed attack presents a severe threat to the emerging AR/VR paradigm, as all existing AR/VR apps can be maliciously turned into an eavesdropper to uncover users' privacy.
- We thoroughly investigate the relationships between the vital signs (i.e., breathing and heartbeat patterns) with facial vibrations. We further design a deep learning model based on a recurrent neural network to reconstruct fine-grained vital sign signals resembling those collected from medical instruments.
- We showcase three representative advanced attacks based on the reconstructed vital signs, including body fat ratio estimation,

user re-identification, and gender recognition. Capitalizing on vital signs with consistent morphological patterns, our attacks can be realized with only a small set (re-identification) or no training data (gender and body fat ratio) from the victim.
- We conduct extensive experiments on three types of mainstream AR/VR headsets, including one low-end cardboard headset and two standalone headsets. The results demonstrate that our attack can reconstruct vital signs with low error rates as well as uncover body fat (less than 4.43%), identity (over 97.83%), and gender of victims (over 93.33%) with high success rates. We also validate that our attack can accurately re-identify users (among 27 participants) under more complicated and practical scenarios, such as across different apps, websites, and longitudinal sessions.

2 ATTACK CONSEQUENCES AND THREAT MODEL

2.1 Attack Consequences

We showcase the attack consequences that could be caused by FaceReader in real-world AR/VR scenarios as follows:

Cyber Tracking. The adversaries can employ the derived physiological status and demographic data to track and profile victims [10, 13, 35]. For example, by correlating vital sign biometrics with network attributes, Internet sessions, and IP addresses, adversaries can pinpoint the geo-location of victims. In addition, once the identity is detected, the victim's anonymous post, message, and app usage history can be compiled to construct a comprehensive profile, shedding light on his/her longitude behaviors and lifestyle. Additionally, private data such as gender and physiological attributes (e.g., breathing patterns, heartbeats, body fat) revealed with our attack further enrich this profile.

Discrimination. The adversaries can leverage the leaked demographic and physiological information via *FaceReader* to perform discrimination against specific individuals or groups. Such discrimination can be reflected in virtual social media platforms (e.g., Horizon Worlds [26], Sensorium Galaxy [33]) and the prices of specific products. For instance, well-targeted hateful and offensive comments can be customized by adversaries on virtual social media platforms based on the demographic information derived from the victims. Moreover, medical insurance companies can make price discrimination [29] by charging different prices of medical insurance after they derive victims' physiological status via *FaceReader*.

Cyberbullying. Cyberbullying denotes using online platforms to harass specific individuals or groups. In recent studies [30, 34], cyber-bullying that correlates with humans' physiological status has been widely reported, where body overweight and underweight have become important factors. With *FaceReader*, the adversaries can derive victims' body fat ratio and perform cyberbullying against overweight or underweight people on AR/VR platforms. Victims who experience such cyberbullying could endure severe emotional distress, such as anxiety, depression, and low self-esteem, leading to a decline in mental health and overall well-being.

2.2 Threat Model

The Adversary. The adversaries could be some enterprises (e.g., AR/VR companies, advertisement agencies, and healthcare vendors) that intend to acquire vital signs and the gender/identity/body fat ratio of users. Such information allows the enterprise to learn important user statistics for product promotion, internal/external collaboration, talent hiring, cross-enterprise business, and even sell the data for benefits. In addition, the adversaries could be a malicious actor (e.g., an application developer or an employee) who has the opportunity to access the motion sensor data. As motion sensors are accessible for AR/VR apps, our attack can be launched at a large scale by reusing existing non-malicious AR/VR apps. The attacker can also acquire the victims' motion sensor data by hosting an innocuous website (e.g., a forum or an online game), which obtains the motion sensor data when users visit the website.

The Adversary's Capability. We assume the adversary can acquire motion sensor data from AR/VR headsets through an AR/VR app or website. The app/website collects motion sensor data in the background and sends it to the adversary for vital sign reconstruction and advanced sensitive information derivation. Particularly, most commercial AR/VR headsets (e.g., Meta Quest, HTC Vive Pro Eye, Valve Index) come equipped with a three-axis accelerometer and a gyroscope, which are utilized to track the user's head motion for motion simulation in the virtual environment. Since motion tracking is an essential feature of AR/VR, accessing the motion sensor typically does not necessitate any user permissions. We confirm the feasibility of stealthy motion sensor data collection by building apps on two mainstream AR/VR programming platforms (Oculus SDK [25], OpenVR SDK [38]) and a website based on WebXR Device API [42]. Our study involves three different headsets (i.e., Meta Quest 1 & 2 and HTC Vive). Particularly, we develop an AR/VR app using the Oculus SDK and successfully employ the function `ovr_GetTrackingState()` to record accelerometer and gyroscope readings from Meta Quest 1 & 2 without requesting user permission. A similar app is implemented with OpenVR, which supports the HTC Vive headset as well as various other headsets (e.g., those manufactured by Valve, Lenovo, and Microsoft). Our app leverages `GetRawTrackedDevicesPoses()` to gather motion sensor data in the background, also without user permission. Furthermore, we verify that motion sensor data can be stealthily collected from all three headsets by hosting a website created with the WebXR Device API, which enables users to interact with the website via their headsets. Our website first requests users' permission to create a session to present immersive content and then utilizes the function `getPose()` to collect motion sensor data in the background without any further permissions.

Attack Scenarios. We study the following four attack scenarios based on vital-sign-induced facial vibrations collected with AR/VR motion sensors and elaborate on their training requirements:

Direct attack: vital sign reconstruction. The adversary can build a deep-learning model to reconstruct high-quality vital signs from the victim's motion sensor data (i.e., facial vibrations). A pre-trained model can be built by correlating the motion sensor data of other people (not including the victim) with the ground-truth breathing and heartbeat signals (e.g., collected using respiration belts and PPG sensors). The adversary can execute a real-time attack leveraging the pre-trained model. Additionally, the adversary may leverage the victim's motion sensor data (without the ground truth) collected in the inference phase to perform adaptive training, which can enhance the performance of the pre-trained model.

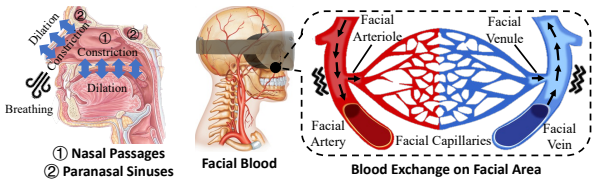


Figure 1: Facial vibrations of human's respiration and heartbeat-related blood exchange process.

Advanced attack: gender recognition. The adversary constructs a deep-learning-based model to discern victims' gender, which can be exploited to identify victims of a particular gender for intrusive or detrimental purposes (e.g., promoting cosmetics for females or action games for males). To train the model, the adversary can utilize motion sensor data and gender labels gathered from other individuals. Once the victim's motion sensor data is acquired (without the gender label), the adversary can either employ the data for adaptive training or initiate a real-time attack without adaptation.

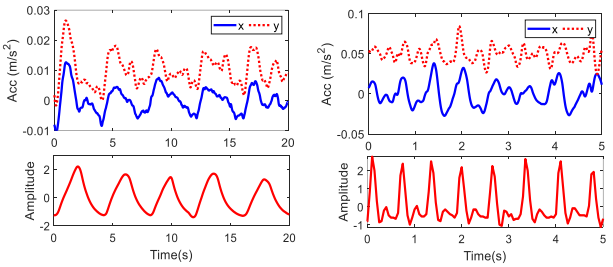
Advanced attack: user re-identification. In this scenario, the adversary aims to identify and track users across various sources of motion sensor data, including but not limited to different AR/VR apps, websites, and sessions. This could enable the adversary to link the users' information from different apps and websites (e.g., game reviews and photos on multiple websites) to discern the user's personalities and even real-world identities. In addition, longitudinal attacks across sessions of using AR/VR devices allow the attacker to recognize the user's interests and preferences. To construct a re-identification model, the adversary trains a deep learning model using the victim's motion sensor data and the user ID (i.e., label) from a single source (e.g., *Source 1*). The profile (deep learning model) only needs to be constructed once based on data from a single source, and the adversary can use this profile to track the user on other sources (e.g., apps, websites, and sessions). It is crucial to note that the user ID could be source-specific and may vary across apps, websites, and sessions, meaning that it is unrealistic to find the same user based on the ID across sources. Given new motion sensor data from a different source (e.g., *Source 2*), the model determines whether the sensor data originates from one of the profiled users with a specific ID in *Source 1*.

Advanced attack: body fat ratio estimation. The adversary may infer the body fat ratio from the victim's reconstructed vital signs. Body fat typically impacts users' cardiovascular systems by clogging arteries and increasing resistance to breathing [4, 8], which allows for deriving body fat information from vital signs. Specifically, the adversary can develop a regression model to correlate the vital sign data with the ground-truth body fat ratio (e.g., obtained using body fat scales [32]) collected from other people. Similar to the direct attack, the adversary can either employ adaptive training to enhance the derivation performance or initiate a real-time attack, without using the victim's body fat information for training.

3 ATTACK OVERVIEW

3.1 Capturing Vital-sign-induced Facial Vibrations via AR/VR Motion Sensors

Kinetic of Breathing and Heartbeat. Human respiration and heartbeat generate subtle facial vibrations on the face surface. In particular, during respiration, inhalation and exhalation cause periodic airflow into and out of the nasal cavities, encompassing nasal passages and sinuses, situated at the roof of the mouth. As illustrated in Figure 1, the process of respiration results in periodic dilation of the human nose. Due to the close contact between the headset and the face, the contraction and expansion of the nasal cavities induce minuscule vibrations on the headset. In addition to respiratory patterns, the headset can also detect vibrations corresponding to heartbeats. The human face comprises a complex network of blood vessels, including the facial artery and its subsidiary branches, such as arterioles and venules, as depicted in Figure 1. In each heartbeat cycle, the heart propels blood through these vessels to the facial tissues, generating minute facial vibrations that can be captured by the AR/VR headset.



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

Figure 2: Comparisons of facial vibrations from AR/VR accelerometer (i.e., x- and y-axis) and corresponding respiration/heartbeat signals captured by medical sensors (i.e., a head-mounted PPG sensor and a respiration belt).

Capturing Facial Vibrations. We conducted two experiments to separately validate the feasibility of detecting facial vibrations induced by breathing and heartbeat. Our first experiment focuses on analyzing the patterns of accelerometer readings from a Meta Quest headset and a NeuLog NUL-236 respiration monitoring belt, with a human participant wearing the headset and the monitoring belt around the waist. The sampling rates are set at 1000Hz for the motion sensors on the Meta Quest and 100Hz for the respiration belt. We compare the patterns of the motion sensor data and the respiration patterns in Figure 2(a). Note that we isolate the sensor data related to breathing-induced facial vibrations using signal separation techniques, which will be discussed further in Section 4.4. We observe similar trends and periodicity of breathing patterns in the facial vibrations and the reference respiration data. In our second experiment, we instruct the participant to wear the headset and attach a face-mounted PPG sensor with a sampling rate of 100Hz. The comparison between heartbeat-related facial vibrations (extracted using techniques that will be described in Section 4.4) and PPG signals is presented in Figure 2(b). We find that heartbeat-induced facial vibrations and heartbeat waveforms display similar morphological patterns to some degree, although the facial vibrations are noisier. The study demonstrates the potential of capturing heartbeat patterns using accelerometers.

3.2 Advantages of Exploiting Facial Vibrations

Our attack, based on vital-sign-induced facial vibrations, presents distinct advantages over prior works in the following perspectives.

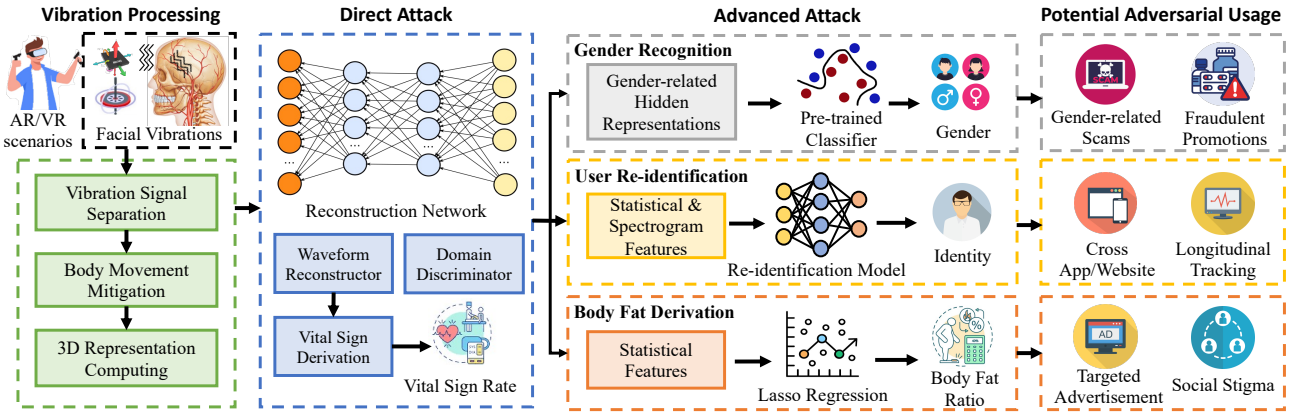


Figure 3: Overview of *FaceReader* for the direct attack on vital sign waveform reconstruction and the advanced attacks for gender recognition, user re-identification, and body fat derivation.

(1) Employing unrestricted sensors. Our attack solely requires data from unrestricted motion sensors, which are accessible by any AR/VR apps and websites. Although some headsets are equipped with PPG/ECG sensors (e.g., heartbeat sensor on HP Reverb G2 Omnicept VR headset [17]), these are considered restricted sensors and necessitate explicit user permissions for access. *(2) Unobtrusive to victims.* Our attack targets spontaneous breathing and heartbeat activities that are naturally produced by the human body, and it is unobtrusive to victims. Differently, prior privacy attacks assume the victim is performing active (obtrusive) interactions, such as speaking [1, 2, 18] and conducting activities [19, 23, 41]. *(3) Enhanced reliability and data-efficiency.* Our attack leverages vital sign signals that exhibit high consistency across different application contexts and scenarios, facilitating sensitive information extraction with significantly less (for re-identification) or no training data (for body fat and gender recognition). In contrast, prior works based on speeches and activities demand extensive training to remove the phonemic and activity variations.

3.3 System Flow

FaceReader consists of three major modules, including *Vibration Processing*, *Direct attack* and *Advanced Attack*. The overview of the attack system is shown in Figure 3.

Vibration Processing. Our vibration processing first separates the 3-axis accelerometer and gyroscope readings into breathing-induced and heartbeat-induced facial vibrations. Then we employ two band-pass filters with cut-off frequencies to extract breathing- and heartbeat-related facial vibrations. A body movement mitigation scheme based on an adaptive filter is then applied to track the vibration signals and dynamically cancel artifacts of human motions. With the denoised vibrations, our attack extracts 3D acceleration, speed, and displacement from breathing- and heartbeat-related vibrations for vital sign waveform reconstruction.

Direct Attack. We design a waveform reconstruction network to realize the proposed direct attack. Specifically, the reconstruction network leverages Long Short-Term Memory (LSTM) units and a self-attention mechanism to correlate the sequential characteristics of facial vibrations with the ground truth (i.e., PPG signals and breathing patterns). The sequential features are utilized by a waveform reconstructor to derive fine-grained breathing and heartbeat waveforms. Additionally, we design a domain discriminator for adaptive training, which improves the reconstruction precision when applied to the victim without the ground truth. Finally, the fine-grained reconstructed vital sign waveforms will be leveraged by adversaries to derive victims' respiration and heartbeat rates.

Advanced Attack. Based on the reconstructed vital sign waveforms, we design a suite of task-specific schemes to perform *Gender Recognition*, *User Re-identification*, and *Body Fat Ratio Estimation*. *(1) Gender Recognition:* We design a hidden representation extractor to extract features from heartbeat and breathing patterns. A light-weight classifier is designed for gender detection without requiring a gender label from the victim. A domain discriminator is built to enhance the generalizability of the model. *(2) User Re-identification.* Our user re-identification scheme includes a statistical feature extractor and a spectrogram feature extractor, which respectively extract unique time and frequency-domain representations from the reconstructed waveforms. We further develop a deep-learning-based model that combines the two types of features to accurately reveal the user's identity. *(3) Body Fat Ratio Estimation.* We design a Lasso-regression-based model to derive body fat ratio leveraging statistical features of the reconstructed vital sign patterns. Our body fat ratio estimation scheme is aligned with prior works on using PPG/ECG signals to derive body fat information [4, 8].

4 ATTACK DESIGN

4.1 3D Representation Computation

Given the particularity of lung capacity, blood flow, and face structure of different people, minute facial vibrations corresponding to human respiration and cardiac cycle show physiological characteristics of different perspectives. For instance, the speed measurements of facial vibration signals could reflect the resistance level of respiration. The displacement measurements of facial vibration cycles may also correlate with the face structure. Based on such considerations, 3D representations of facial vibrations, such as 3D speed and displacement, could depict these geometric characteristics of the human's head as well as the facial vibration magnitude during each cycle of respiration/heartbeat. These representations allow the attacker to derive sensitive physiological information. Particularly,

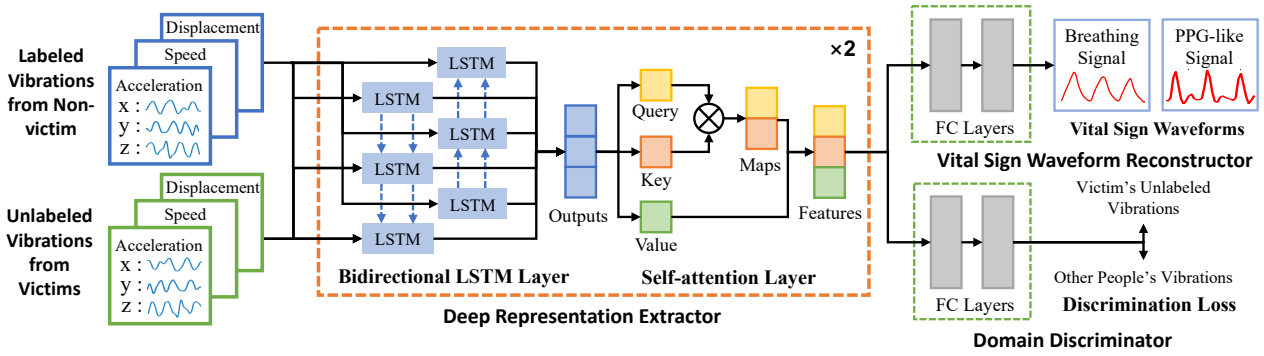


Figure 4: Overview of the fine-grained vital sign waveform reconstruction model.

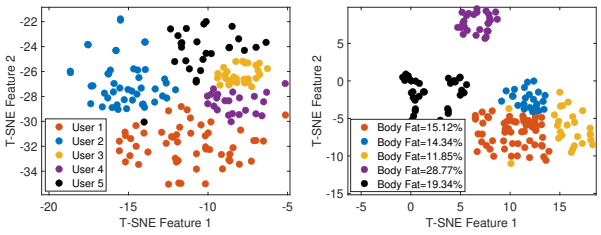
we calculate the 3D speed and displacement from facial vibration signals leveraging first-order and second-order integration from the 3-axis readings collected from unrestricted AR/VR motion sensors. Our attack then leverages these 3D accelerations, speed, and displacement as inputs for the direct attack.

4.2 Direct Attack: Vital Sign Waveform Reconstruction

In Section 3.1, we show that facial vibrations captured by AR/VR motion sensors suffer severe distortions caused by hardware noises compared to ground-truth vital signs. These hardware noises have orders of amplitudes than those of minute facial vibrations induced by human vital signs. In such cases, the fine-grained morphological features of vital signs, such as systolic and diastolic peaks in each heartbeat cycle, are significantly distorted and hard to detect. Traditional signal processing methods usually fail in recovering such patterns since they rely on the assumption that the signals have a sufficiently higher signal-to-noise ratio (SNR). Additionally, while some AR/VR headsets (e.g., HP Reverb G2 Omnicept VR headset [17]) have been equipped with heartbeat sensors, accessing these sensors is typically restricted by the operating system. Under this circumstance, the attackers could not acquire the ground truth vital sign patterns for training the vital sign reconstruction model. Motivated by prior works [45, 46] using deep learning techniques to boost SNR in wireless sensing, we develop a signal reconstruction model for adversaries to remove hardware noise and reconstruct fine-grained vital signs. The model learns to correlate facial vibrations with PPG signals (i.e., ground truth). We further develop a domain adaptation approach to adapt the model parameters based on unlabeled motion sensor data of the victim, as in practical scenarios, the ground truth is not available.

Fine-grained Vital Sign Reconstruction Network. Since respiration and heartbeat waveforms manifest strong characteristics of periodicity and consistency, we build the deep representation extractor based on two consequent bidirectional-LSTM layers to expose periodic features from facial vibration signals. Specifically, we set the output units of the two bidirectional LSTM layers as 1024 and 512, respectively. In order to capture the internal dependencies within each waveform segment, particularly in the areas of inhalation/exhalation and systolic/diastolic cycles, we incorporate a self-attention layer for each bidirectional-LSTM layer. For these two self-attention layers, we make the Query (Q), Key (K), and Value (V) equal to the output dimension of their connected LSTM layers. The detailed model architecture is depicted in Figure 4. Specifically, the reconstruction model takes 6 channels corresponding to the readings from the 3-axis accelerometer and gyroscope as inputs. For each channel, it includes several segments $x_1, x_2, ..., x_n$, which are segmented from the motion sensor readings $\mathcal{X}$. The deep representation extractor $\mathcal{F}(\cdot)$ consists of two blocks of bidirectional-LSTM layers and self-attention layers, which extract deep representations from motion sensor readings. Finally, the extracted deep representation will pass through a waveform reconstructor $\mathcal{D}(\cdot)$, which includes two fully connected layers, to reconstruct fine-grained respiration and heartbeat patterns.

Domain Discriminator for Adaptive Training. The vital sign reconstruction model may work well on people with ground-truth vital sign signals for training. However, the performance may degrade when being directly applied to a new user (victim). Training the model on a large dataset may improve the generalizability, but it will introduce significant training costs, making the attack difficult to launch in more practical scenarios. To enable the waveform reconstructor $\mathcal{D}(\cdot)$ applicable to new victims, we develop a domain adaptation scheme, which effectively transfers the prior knowledge in the pre-trained vital sign reconstructor to unknown victims. As a potent machine learning technique, domain adaptation empowers models to extend their generalizability from a source domain (i.e., data from recruited users) to another distinct target domain (i.e., data from victims). The essence of domain adaptation is to align the representations of the source and target domain, making the model focus on the shared and domain-independent features, thus enhancing the model's performance on the target domain. Specifically, we achieve this by designing a domain discriminator $\mathcal{Z}(\cdot)$ with two fully connected layers, which is depicted in Figure 4. The domain discriminator $\mathcal{Z}(\cdot)$ takes the representations from the deep representation extractor $\mathcal{F}(\cdot)$ as inputs and outputs the domain label (i.e., the victims or the other people recruited by the attackers). Then we employ a generative adversarial loss [12] on the domain discriminator $\mathcal{Z}(\cdot)$, which guides the deep representation extractor $\mathcal{F}(\cdot)$ to learn the user-independent features. Particularly, the idea is to apply a negative factor to the loss function for updating the parameters of the domain discriminator $\mathcal{Z}(\cdot)$, so that the trainable parameters of the deep representation extractor can be updated through the negative loss until it can "confuse" the domain discriminator. After involving the domain adaption scheme, the deep representation extractor $\mathcal{F}(\cdot)$ is able to extract user-independent



(a) T-SNE on user identities (b) T-SNE on body fat ratios

Figure 5: T-SNE of user-specific statistical features showing distinctive clusters based on different user identities and body fat ratios.

representations to the greatest possible extent, thus maximizing the generalization performance of the pre-trained reconstruction model on the victims.

Model Training. During the training phase of the deep representation extractor $\mathcal{F}(\cdot)$, the waveform reconstructor $\mathcal{D}(\cdot)$ and the domain discriminator $\mathcal{Z}(\cdot)$, the ground truth signals of respiration belt and PPG sensors (e.g., collected from other people recruited by the attackers) are segmented to $y_1, y_2, ..., y_n$ corresponding to facial vibration segments $x_1, x_2, ..., x_n$. Given the sensor readings and their corresponding ground truth signals captured by medical sensors, we optimize the parameters θ, ω, and γ of the deep representation extractor $\mathcal{F}_\theta(\cdot)$, waveform reconstructor $\mathcal{D}_\omega(\cdot)$ and domain discriminator $\mathcal{Z}_\gamma(\cdot)$, which can be described as:

$$\mathcal{L}_R = \sum_{i=1}^{n} \mathcal{L}_{MSE}(\mathcal{D}_\omega(\mathcal{F}_\theta(x_i)), y_i), \mathcal{L}_D = \sum_{i=1}^{n} \mathcal{L}(\mathcal{Z}_\gamma(\mathcal{F}_\theta(x_i)), d_i),$$
$$\arg\min_{\theta} \mathcal{L}_R - \lambda\mathcal{L}_D, \arg\min_{\omega} \mathcal{L}_R, \arg\min_{\gamma} \mathcal{L}_D, \quad (1)$$

where d_i denotes the domain label (e.g., the victim individual or users recruited by the attacker) of the facial vibration segment x_i, $\mathcal{L}_{MSE}$ and $\mathcal{L}$ refer to the Mean Square Error (MSE) and cross-entropy loss, respectively. λ works as the negative factor described previously for balancing the trade-off between the transferability of deep representations and the distinctiveness of different domains, which is set to be 0.1 empirically.

Respiration and Heartbeat Rate Derivation. To derive the respiration and heartbeat rates, time-frequency analysis is applied to the reconstructed vital sign waveforms. Specifically, we apply a Hann Window on the reconstructed waveforms to highlight the responses of dominating frequencies and utilize Short-Time Fourier Transform (STFT) on these windowed respiration and heartbeat waveforms. Subsequently, we select the highest frequency response within 0.1Hz ~ 0.5Hz to derive the victim's respiration rate and within 0.8Hz ~ 3.0Hz to expose the victim's heartbeat rate.

4.3 Advanced Attack: High-level Private Information Derivation

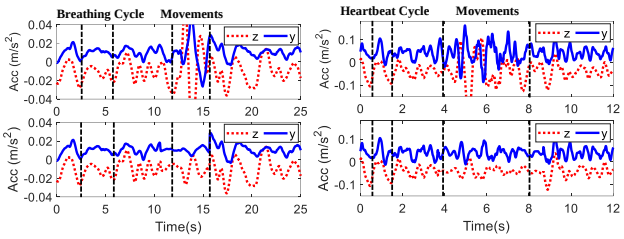
To realize high-level private information derivation, we design a suite of schemes to extract biometrics and biomarkers embedded in the reconstructed vital sign waveforms. **Gender Recognition.** We design a gender-related hidden feature and a lightweight pre-trained gender classifier, which can be effectively and adaptively applied to entirely different user groups. Previous studies have shown the potential of inferring genders based on PPG signals [5, 9, 28], and we show that gender derivation is also feasible on the reconstructed vital signs from facial vibrations. Our scheme extracts hidden representations from the penultimate fully connected layer of the fine-grained vital sign waveform reconstruction network described in Section 4.2, which connect facial vibrations to PPG-like signals, where the representations contain rich biometric characteristics from both facial vibrations and PPG signals. By leveraging extracted gender-related hidden representations, we build a lightweight classifier based on the Support Vector Machine (SVM) with the Radical Basis Function (RBF) kernel to infer the victim's gender information. Specifically, gender-related features are extracted from several known male and female users (e.g., friends of the attacker) and utilized to pre-train the SVM-based classifier. During further evaluations, the classifier is demonstrated to be effective in recognizing victims' gender information that does not exist in the training group.

User Re-identification. We design a statistical feature extractor and a spectrogram feature extractor to extract representative features from the time and frequency domain. A dedicated identity derivation network is developed to re-identify users by combining time-domain statistical features and frequency-domain spectrogram features extracted from reconstructed vital sign waveforms.

Statistical Features. Statistical features of vital signs contain rich identity-specific physiological characteristics. For instance, the maximum value of the reconstructed breathing and heartbeat patterns could indicate the highest pressure of a human's breathing and the systolic peak of a human's heartbeat, respectively. To extract these user-specific biomarkers from reconstructed vital sign series and facial vibrations, we apply a sliding window on the short-time reconstructed segments $\mathcal{D}_\omega(\mathcal{Z}_\theta(x_1)), \mathcal{D}_\omega(\mathcal{Z}_\theta(x_2)), ..., \mathcal{D}_\omega(\mathcal{Z}_\theta(x_n))$ and extract 13 types of statistical features, including maximum, minimum, range, mean, variance, root mean square, median, interquartile range, mean crossing rate, skewness, kurtosis, entropy and signal power. Figure 5(a) shows that the statistical features of 5 different users (with similar body fat ratio) have distinctive T-Distributed Stochastic Neighbor Embedding (T-SNE) clusters, which indicate the effectiveness of the statistical features.

Spectrogram Features. The reconstructed respiration/heartbeat waveforms produce frequency spectrograms that contain user-specific features. The user-specific facial structures and components produce unique vibration patterns associated with breathing and heartbeat. Such properties could be utilized to differentiate users. To extract user-specific spectrogram features from reconstructed signals, Short-Time Fourier Transform (STFT) is applied to each reconstructed segment. These user-specific spectrogram features will be utilized by the attackers to re-identify different users.

Re-identification Model Design. We further explore realizing precise user re-identification leveraging the aforementioned features by developing a dedicated deep learning model, which comprises a deep representation extractor and an identity classifier. The deep representation extractor takes four features as inputs, including statistical/spectrogram features of facial vibrations $x_1, x_2, ..., x_n$ and statistical/spectrogram features of reconstructed vital sign signals $\mathcal{D}_\omega(\mathcal{Z}_\theta(x_1)), \mathcal{D}_\omega(\mathcal{Z}_\theta(x_2)), ..., \mathcal{D}_\omega(\mathcal{Z}_\theta(x_n))$. For the statistical feature inputs, we employ a representation extractor with two sequential 2D convolutional layers. 13 types of statistical features from



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

Figure 6: Comparison of facial vibration signals before and after adaptive-filter-based body movement mitigation corresponding to human's respiration and heartbeat.

each time window are vertically stacked and fed into the representation extractor. For the spectrogram inputs, we design another representation extractor including two 2D convolutional layers and two LSTM layers to extract deep representations from frequency-domain spectrograms. Subsequently, the deep representations of four different input features are concatenated as a larger representation for differentiating users. To train the representation extractor and the classifier, we apply Triplet Loss [16] as the loss function, which stretches the input features into a larger space, allowing extracting effective representations with a smaller amount of training data. The training process can be described as:

$$\arg\min_{\theta} \sum_{b,p,n} \max\left(\left\|\mathcal{F}_\theta(b) - \mathcal{F}_\theta(p)\right\|_2 - \left\|\mathcal{F}_\theta(b) - \mathcal{F}_\theta(n)\right\|_2 + \alpha, 0\right), \quad s.t.\ y_b = y_p \neq y_n, \tag{2}$$

where $\mathcal{F}$, θ, and α refer to the representation extractor, trainable parameters, and the margin between positive and negative pairs (we set it as 0.1 empirically). b, p, n, and y_i represent baseline inputs, positive inputs, negative inputs, and corresponding labels, respectively. After deep representation extraction, the concatenated representation will be fed into a deep-learning-based classifier, which consists of three fully connected layers and a softmax layer, to predict the user's identity. During the training phase of the identity classifier, we apply cross-entropy loss as the loss function.

Body Fat Ratio Derivation. To derive the body fat ratio from different users, we leverage the statistical feature extractor from user re-identification to extract representative and unique features and a regression-based derivation model to realize precise body fat ratio prediction. We show examples of T-SNE from 5 different users in Figure 5. In particular, 3 users have relatively lower and similar body fat ratios (i.e., 11.85%, 14.34% and 15.12%) and the other 2 users have higher body fat ratios (i.e., 19.34% and 28.77%). The clusters indicate that user-specific statistical features are distinguishable for different body fat ratios. Based on user-specific statistical features, we design a regression-based framework to further derive the approximate value of victims' body fat ratio. Specifically, we apply a Lasso-regression-based approach to derive victims' body fat ratio based on statistical features of reconstructed breathing and heartbeat patterns. The regression process can be described as:

$$\arg\min_{\mathcal{W}} \sum_{i=1}^{n} (y_i - y_g) = \sum_{i=1}^{n} (y_g - \sum_{j=0}^{p} \mathcal{W}_j \cdot x_{i,j}) + \lambda \cdot \sum_{j=0}^{p} |\mathcal{W}_j|, \tag{3}$$

where $\mathcal{W}$, y_i, y_g, $x_{i,j}$ refer to the regression weights of each input feature, the prediction result, the ground truth of human's body fat ratio and the j^{th} input feature of i^{th} facial vibration segment, respectively. n and p represent the total number of input samples and the number of features we involve in our regression-based model (e.g., 13 different statistical features). λ is denoted to be the weights of the Lasso penalty term to help control over-fitting, which is set to be 0.5 empirically in this task.

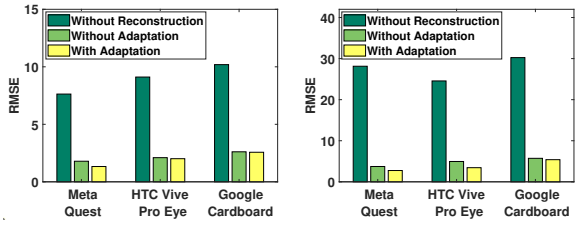
4.4 Activity Mitigation

Vibration Signal Separation. Respiration and heartbeat patterns usually have distinctive and independent frequency ranges of facial vibrations. Human usually breathes 12 ∼ 16 cycles and the heart pumps 48 ∼ 180 times per minute. Accordingly, two level-2 Butterworth band-pass filters with cut-off frequency ranges of 0.1Hz ∼ 0.5Hz and 0.8Hz ∼ 3.0Hz are utilized to separate vibrations induced by human's breathing and heartbeat, respectively.

Body Movement Mitigation. During practical usage of AR/VR headsets, users interact with the devices through different types of body movements. Thus, the built-in motion sensors on AR/VR headsets can pick up various and unpredictable artifacts associated with human motions, such as spontaneous head movements and non-spontaneous body shaking. Since body motions in AR/VR scenarios typically share similar frequency components with human respiration and heartbeat, the artifacts induced by body motions cannot be filtered through fundamental band-pass filters. To mitigate such artifacts, we design a generalized scheme based on Short-Time-Energy (STE) computation and adaptive filtering. For each segment of readings from the 3-axis accelerometer and gyroscope, our method continuously computes the total energy with a specific time window. This energy computation is compared with a predefined power threshold, which detects the existence of body motions in the signals under test. Once body motions are detected, an adaptive-filter-based method is employed to mitigate the artifacts. We realize the adaptive filter by solving the following optimization problem to determine the adaptive weight vector $\mathcal{W}$:

$$\begin{aligned} \arg\min_{\mathcal{W}} \sum_{t\in\mathcal{T}} \mathcal{E}(t) &= \sum_{t\in\mathcal{T}} \mathcal{D}_{KL}(\mathcal{R}(t)||\mathcal{X}(t)) = \sum_{t\in\mathcal{T}} \mathcal{R}(t)\, log\frac{\mathcal{R}(t)}{\mathcal{X}(t)} \\ \text{s.t.} \sum_{t\in\mathcal{T}} \mathcal{W}(t) &= \alpha \cdot \sum_{t\in\mathcal{T}} \mathcal{W}(t) + \mu \cdot \sum_{t\in\mathcal{T}} \mathcal{E}(t)\cdot\widetilde{\mathcal{X}}(t), \\ \sum_{t\in\mathcal{T}} \mathcal{X}(t) &= \sum_{t\in\mathcal{T}} \mathcal{W}(t)\cdot\widetilde{\mathcal{X}}(t), \sum_{t\in\mathcal{T}} \widetilde{\mathcal{X}}(t)^2 \leq \rho_{power}, \end{aligned} \tag{4}$$

where $\widetilde{\mathcal{X}}$, $\mathcal{X}$, $\mathcal{E}$, and $\mathcal{R}$ represent the signals before mitigation, the signals after mitigation, error function, and the reference signal for the adaptive filter, respectively. The signal index, leakage parameter, and step size are denoted by $\mathcal{T}$, α, and μ, respectively, and are referred to as hyper-parameters in this optimization problem. For each segment of readings from the 3-axis accelerometer and gyroscope, we select a specific value ρ (0.2 empirically) as the threshold to determine whether the signal segments need to be adaptively filtered or not. Since Kullback-Leibler Divergence [21] gives a reliable quantification on the difference of distributions of two signal series [20], we select it as the error measurement between unfiltered and reference signals, and to optimize the weights of the adaptive filter. Figure 6(a) and Figure 6(b) demonstrate the



(a) Breathing pattern reconstruction (b) Heartbeat pattern reconstruction

Figure 7: Attack performance of breathing and heartbeat reconstruction with/without victim's data for adaptive training compared with raw facial vibration signals.

effects before and after body motion mitigation. We can observe that the artifacts are significantly mitigated after passing our proposed scheme, which further manifests the effectiveness of our adaptive-filter-based approach.

5 EVALUATION I: PERFORMANCE OF VITAL SIGN RECONSTRUCTION

5.1 Experimental Setup

AR/VR Devices. *FaceReader* is validated upon two standalone AR/VR headsets (i.e., Meta Quest and HTC Vive Pro Eye) and one low-cost smartphone-based headset (i.e., Google Cardboard with Samsung Galaxy S6 smartphone). The motion sensor module used by HTC Vive Pro Eye and Samsung Galaxy S6 is Invensense MPU-6500, which includes an accelerometer/gyroscope with a resolution of 2048/4096/8192/16384 LSB/g and 16.4/32.8/65.5/131 $LSB/\circ/sec$, respectively. Differently, Meta Quest is equipped with a motion sensor board, 330-00193-03 1PASF8K, which is originally designed by Meta. For our proposed attack, we set the sampling rate of motion sensors on Meta Quest, HTC Vive Pro Eye, and Samsung Galaxy S6 to be 1000Hz, 1000Hz, and 203Hz, respectively, which are the highest and most stable sampling rates these headsets can achieve. We build a tool to collect motion sensor data in the background from Meta Quest based on Oculus (Meta) SDK [25]. Our HTC Vive Pro headset is connected to a desktop computer with an Intel E5-2630v4 processor and Nvidia Quadro GV100 Graphics Card running on Windows 10. We use the OpenVR SDK [38] to build the tool to collect data from the headset. For Google Cardboard, we use the Samsung Galaxy S6 smartphone running on Android OS and develop a tool using the Android SDK [14]. Additionally, we use WebXR Device API [42] to build and host a webpage, which can be accessed by all three headsets.

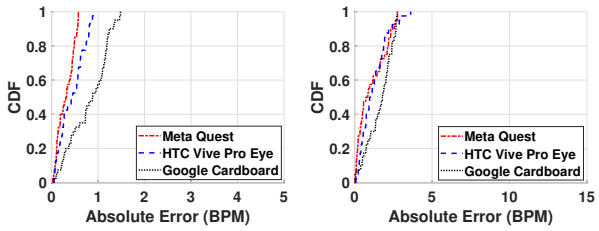
Participants and Data Collection. The data collection corresponding to the direct attack involves 13 users (10 males and 3 females) for Meta Quest, 11 users (8 males and 3 females) for HTC Vive Pro Eye, and 10 users (6 males and 4 females) for Google Cardboard with Samsung Galaxy S6, respectively. To obtain ground truth measurements for respiration and heartbeat patterns, we use the NeuLog NUL-236 Respiration Monitor Belt and NeuLog NUL-208 Photoplethysmography Monitor. The ground truth respiration/heartbeat rates are calculated leveraging the ground truth signals captured from corresponding devices. Specifically, we take turns considering one user as the victim and build the pre-trained vital sign reconstruction model using the labeled data of other users (i.e., motion sensor readings with ground truth vital sign signals). The data collection process has been approved by our university's Institutional Review Board (IRB).

Body Movements During Data Collection. We evaluate our attack under four different types of typical body motions in AR/VR scenarios. *(1) Sitting and Watching a Demo Video.* The user sits in a chair and watches a demo video for 1 minute. During the experiment, the participant remains in a static position and does not perform body movements. *(2) Standing and Watching a Demo Video.* The participant is requested to stand and watch a demo video for 1 minute. During the experiments, the participant may perform some non-spontaneous body movements (e.g., body shaking) compared to the sitting scenario. *(3) Using Controllers to Browse the AR/VR Webpage.* In this scenario, the participant performs spontaneous arm movement similar to arm raising and dropping. While spontaneous motions are involved in this scenario, the motions are arm movements and could not induce direct and significant fluctuations on the motion sensors of the head-mounted AR/VR device. *(4) Walking in the Virtual Environment.* In this scenario, the participant is requested to walk along a 2-meter trajectory within the virtual environment, which lasts for 3 seconds approximately. Walking introduces large-scale spontaneous movements and induces significant changes in AR/VR motion sensor readings when the participant starts or stops walking. All our experiments are conducted under the aforementioned four scenarios with typical non-spontaneous and spontaneous body movements. We apply our designed adaptive-filter-based body movement mitigation scheme to remove the artifacts.

Evaluation Metrics. We use two metrics to quantify the effectiveness of *FaceReader* for the direct attack, including respiration/heartbeat reconstruction and respiration/heartbeat rate derivation. *(1) Root Mean Square Error (RMSE):* We evaluate the performance of the direct attack on vital sign waveform reconstruction through computing RMSE between the reconstructed breathing/heartbeat waveform and patterns captured by medical sensors (e.g., respiration belts and PPG sensors). *(2) Absolute Error (AE):* We utilize AE to evaluate the performance of respiration/heartbeat rate derivation. Specifically, we measure the AEs between the derived rates and the ground truth rates calculated from the head-mounted PPG sensors and the respiration belt.

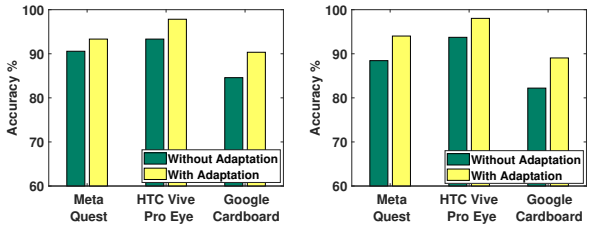
5.2 Evaluation of Breathing and Heartbeat Waveform Reconstruction

We evaluate the performance of the direct attack that reconstructs breathing and heartbeat waveforms on three different types of commodity AR/VR headsets. We show the results in Figure 7. Specifically, for Meta Quest and HTC Vive Pro Eye, the RMSEs between facial vibrations and ground truth breathing signals (without reconstruction) are more than 7.63 and 9.11. For ground truth heartbeat signals, the RMSEs without reconstruction are more than 28.14 and 24.56 for Meta Quest and HTC Vive Pro Eye. After involving respiration reconstruction, the RMSEs between reconstructed breathing waveform and ground truth breathing signals on Meta Quest and HTC Vive Pro Eye can achieve less than 1.33 and 2.01 with motion sensor data from victims (particularly the ground-truth vital signs) for adaptive training, and less than 1.79 and 2.10 without any sensor



(a) Breathing rate derivation (b) Heartbeat rate derivation

Figure 8: Attack performance of breathing/heartbeat rate derivation from reconstructed breathing/heartbeat signals.



(a) Performance of cross-app/website (b) Performance of cross-device

Figure 9: Attack performance of gender recognition on cross-app/website scenario and cross-device scenario.

data from victims. For heartbeat reconstruction, the RMSEs of Meta Quest and HTC Vive Pro Eye are lower than 2.75 and 3.73 with adaptive training, and lower than 3.44 and 4.95 without adaptive training. For the low-end Google Cardboard headset with less sensitive motion sensors (e.g., lower sampling rate), the RMSEs between facial vibrations and reconstructed fine-grained respiration and heartbeat signals are greater than 10.19 and 30.24 without vital sign reconstruction. After involving the breathing and heartbeat reconstruction model, the RMSEs can achieve less than 2.61 and 5.72 without adaptive training, and less than 2.57 and 5.39 with adaptive training. The results demonstrate that the direct attack on breathing and heartbeat waveform reconstruction is effective on different types of commodity AR/VR headsets.

5.3 Breathing/Heartbeat Rate Derivation

Next, we measure the absolute errors (AEs) between our breathing/heartbeat rate measurements of reconstructed signals and ground truth breathing/heartbeat rate captured by medical devices. Particularly, Beats Per Minute (BPM) is utilized to measure the AEs of breathing and heartbeat rate derivation. The performance of three different AR/VR headsets is shown in Figure 8. For Meta Quest, the absolute errors of *FaceReader* on deriving breathing and heartbeat rates are lower than 0.62BPM and 0.92BPM. For HTC Vive Pro Eye, we also achieve absolute errors lower than 0.87BPM and 3.62BPM for breathing and heartbeat rate derivation, respectively. For the low-end Google Cardboard headset with less sensitive motion sensors, the absolute errors of breathing and heartbeat rate derivation are less than 1.50BPM and 2.75BPM. The results confirm the effectiveness of our direct attack for breathing and heartbeat rate derivation on different types of commodity AR/VR headsets.

6 EVALUATION II: PERFORMANCE OF ADVANCED ATTACK

6.1 Experimental Setup

Data Collection. Our dataset includes 27 users (19 males and 8 females) for Meta Quest, 25 users (17 males and 8 females) for HTC Vive Pro Eye, and 15 users (8 males and 7 females) for Google Cardboard with Samsung Galaxy S6, respectively. We involve all four body movements mentioned in Section 5.1 during the data collection and use the dataset to evaluate *FaceReader* for gender recognition and user re-identification. Regarding body fat ratio estimation, we use a RENPHO Smart Body Fat Scale to track the body fat percentage of 10 users for one month. While this device provides only consumer-level body fat estimation, we aim to demonstrate that our proposed attack can achieve the potential derivation of a victim's body fat rate. Particularly, the data collection covers the common body fat percentage ranging from 7.8% to 28.9%. All data collection processes have been approved by our university's Institutional Review Board (IRB).

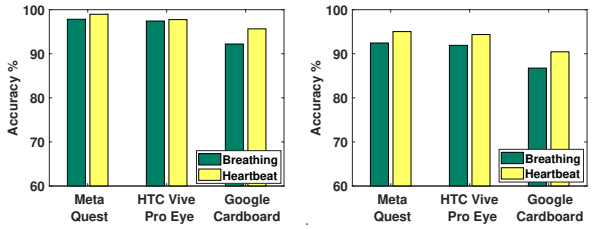
Evaluation Metrics. *(1) Derivation Accuracy.* To evaluate the performance of advanced attacks on gender recognition and user re-identification, we utilize derivation accuracy as a metric. This metric measures the percentage of data samples that are correctly recognized as belonging to the correct gender/user as the ground truth labels. *(2) Absolute Error (AE).* We evaluate the effectiveness of our proposed advanced attack on the task of body fat ratio derivation. Specifically, we measure the AEs between the derived ratio and the ground truth measured from the smart body fat scale.

6.2 Gender Recognition

We validate the performance of *FaceReader* in gender recognition through evaluations conducted in two distinct attack scenarios: cross-app/website attack and cross-device attack. In the cross-app/website attack, the victim utilizes a different AR/VR app/website than the one the attacker employs to collect motion sensor data for building the gender recognition model. In the cross-device attack, the attack on gender recognition is implemented within the same application/website and across the same type of AR/VR devices (i.e., across two Meta Quests).

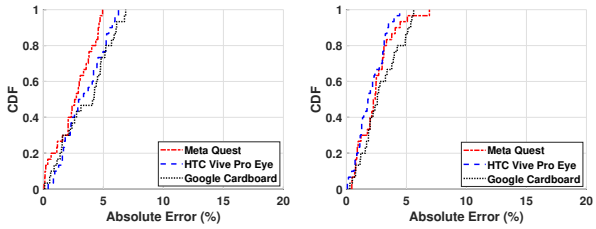
Cross-app/website Attack. We evaluate the gender recognition accuracy under the cross-app/website scenario, and the performance is shown in Figure 9(a). Regarding Meta Quest, *FaceReader*achieves a gender recognition accuracy of 93.33% with adaptive training and over 90.57% accuracy without adaptive training. For HTC Vive Pro Eye, *FaceReader*'s gender recognition also achieves over 97.83% and 93.33% accuracy with and without adaptive training, respectively. In the case of Google Cardboard with less sensitive sensors, the gender recognition accuracy can also surpass 90.34% and 84.58% with and without adaptive training. High gender recognition accuracy from three types of AR/VR headsets demonstrates the effectiveness of gender recognition under the cross-app/website scenario.

Cross-device Attack. We also conduct evaluations on cross-device scenarios for gender recognition. As shown in Figure 9(b), the accuracy of gender recognition for Meta Quest can achieve 88.44% without adaptive training and over 94.02% with adaptive training. For HTC Vive Pro Eye, the accuracy of gender recognition



(a) Performance of cross-app/website (b) Performance of cross-device

Figure 10: Performance of user re-identification on the same app/website and across different apps/websites.



(a) Body fat derivation from breathing (b) Body fat derivation from heartbeat

Figure 11: Performance of body fat percentage derivation from facial vibrations induced by breathing and heartbeat.

can also achieve 98.03% and 93.73% with and without adaptive training. For Google Cardboard with less sensitive motion sensors, the gender recognition accuracy can also reach over 89.05% with adaptive training and 82.21% without adaptive training. The results demonstrate that our advanced attack on gender recognition can realize precise recognition of the victims' gender information across different devices.

6.3 User Re-identification

To validate *FaceReader* on the user re-identification task, we involve two separate evaluations that simulate practical AR/VR scenarios, including cross-app/website and cross-session. Specifically, in the cross-app/website scenario, the victim employs a different AR/VR app or website from the ones attackers use to gather motion sensor data and construct the re-identification model. In the cross-session scenario, testing data from the victims are collected on different days. In this scenario, we simulate practical and longitudinal user tracking by leveraging our designed re-identification approach.

Cross-app/website Attack for User Re-identification. We evaluate the accuracy of user re-identification in a cross-app/website attack scenario. The re-identification accuracy for three types of common AR/VR headsets is depicted in Figure 10(a). From the results, we observe that prominent accuracy is achieved in user re-identification via all three commodity AR/VR headsets. Specifically, the re-identification accuracy surpasses 97.83% and 98.96% on Meta Quest, leveraging reconstructed breathing and heartbeat waveforms. For HTC Vive Pro Eye, *FaceReader*'s accuracy in user re-identification also exceeds 97.43% and 97.75%, utilizing finely reconstructed breathing and heartbeat patterns. Similarly, on Google Cardboard with less sensitive motion sensors, the re-identification accuracy can reach over 92.21% and 95.65%, harnessing reconstructed respiration and heartbeat signals. This high accuracy observed across the three types of common AR/VR headsets underscores the effectiveness of the user re-identification attack design when deployed across AR/VR apps or websites.

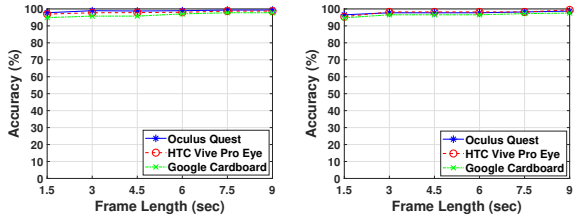
Cross-session Attack for User Re-identification. We also assess the performance of user re-identification in the context of the cross-session attack scenario. Specifically, we set a one-week interval between the training data collection for training the user re-identification model and testing data collection. The performance of longitudinal cross-session user re-identification is depicted in Figure 10(b). The re-identification accuracy on Meta Quest can exceed 92.44% and 95.03%, leveraging reconstructed breathing and heartbeat signals. For HTC Vive Pro Eye, *FaceReader* achieves a re-identification accuracy of over 91.91% and 94.36%, utilizing finely reconstructed respiration and heartbeat patterns. Even for the lower-end Google Cardboard headset, the re-identification accuracy surpasses 86.75% and 90.44% with reconstructed breathing and heartbeat signals. Remarkably, despite the one-week gap between the collection of training and testing data from AR/VR motion sensors, *FaceReader* maintains a high level of accuracy in user re-identification across distinct AR/VR usage sessions. This underscores the feasibility of utilizing *FaceReader* for longitudinal user tracking in practical AR/VR scenarios.

6.4 Body Fat Percentage Derivation

To evaluate the performance of the advanced attack on deriving the victim's body fat ratio, we measure the absolute errors between the Lasso regression results and the ground truth body fat ratio measured via the smart body fat scale. Specifically, after constructing the regression model using vital-sign-induced facial vibrations from existing users, we directly apply the regression model to derive the body fat ratio of victims. The derived measurements are presented in Figure 11, and the absolute errors are found to be under 4.92% and 6.91% for Meta Quest, utilizing features from reconstructed breathing and heartbeat patterns. For HTC Vive Pro, the absolute errors can also be less than 6.25% and 4.43% with reconstructed breathing and heartbeat patterns. Similarly, for the lower-end Google Cardboard headset with less sensitive motion sensors, the absolute errors resulting from our advanced attack on body fat ratio derivation are less than 6.85% and 5.59% using reconstructed breathing and heartbeat patterns. The results effectively demonstrate the success of our advanced attack in deriving the body fat ratio across different types of commodity AR/VR headsets.

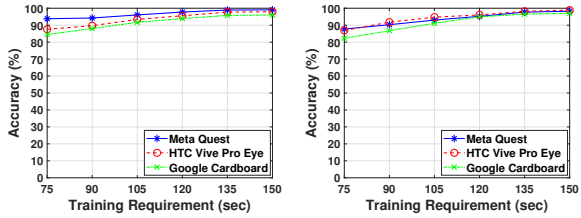
6.5 Evaluation on Attack Robustness

Different Frame Lengths of Training Samples. From the adversary's perspective, if *FaceReader* can achieve high re-identification and gender recognition accuracy with a shorter frame of training samples (e.g., less than 3 seconds), the adversary can derive the victim's private information with lower attack cost (e.g., shorter data collection time). Consequently, we evaluate the impact of varying frame lengths of training samples for user re-identification and gender recognition tasks using heartbeat-induced facial vibration signals as inputs, as illustrated in Figure 12. It is evident from the results that our proposed attack can achieve over 97.75%, 96.94%, and 94.90% accuracy in user re-identification, as well as more than 96.53%, 95.43%, and 94.77% accuracy in gender recognition for Meta



(a) User Re-identification (b) Gender Recognition

Figure 12: User re-identification and gender recognition accuracy across different frame lengths of training samples.



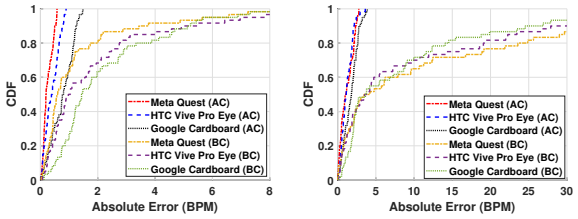
(a) Performance of identity derivation (b) Performance of gender detection

Figure 13: User re-identification and gender recognition accuracy with different training requirements (average data collection time for each user).

Quest, HTC Vive Pro Eye, and Google Cardboard, respectively, with input signals shorter than 1.5 seconds. This substantiates that *FaceReader* could be readily applicable in practical scenarios.

Study of Training Requirements. To study the training requirements of *FaceReader*, we evaluate the performance of user re-identification and gender recognition tasks with different numbers of data samples for training. Specifically, we collect motion sensor data from each user over different time durations. The results of re-identification and gender recognition accuracy are shown in Figure 13. From the results, we observe that *FaceReader* consistently maintains accuracy levels of more than 93.82%, 87.62%, and 84.50% on user re-identification, as well as over 87.75%, 86.88%, and 94.77% on gender recognition for Meta Quest, HTC Vive Pro Eye, and Google Cardboard, respectively, even with minimal training requirements (e.g., less than 75 seconds of data collection per user). The notable outcomes of achieving high accuracy in both re-identification and gender recognition, despite minimal training requirements, indicate that *FaceReader* can be effortlessly implemented in practical scenarios with significant implications. This underscores the importance of AR/VR users giving ample attention to the potential implications.

Impact of Body Movement Mitigation. We further conduct a case study to specifically validate the effectiveness of body movement mitigation scheme under strong body movements. We consider two additional movements in practical AR/VR scenarios: *(1) Head Movements.* Participants take turns rotating their heads from left to right and then from right to left. Data collection for this scenario lasts for 1 minute. *(2) Speaking Random Digits.* In this scenario, participants select two random digits from the TIDigits corpus [31], each of which is spoken at a random time slot during the 1-minute data collection. Speaking introduces entirely different responses that distribute in higher frequency ranges (e.g., ≥ 100Hz)



(a) Breathing rate derivation (b) Heartbeat rate derivation

Figure 14: Performance of breathing/heartbeat rate derivation before (BC) and after (AC) body movement mitigation.

compared to common body movements. Specifically, we compare the performance under a total of six different AR/VR scenarios before and after our proposed body movement mitigation scheme. The results are shown in Figure 14. When compared to the baseline (BC) without motion artifact mitigation, the absolute errors of breathing rate derivation are improved from 8.13BPM, 8.75BPM, and 10.43BPM to 0.57BPM, 0.89BPM, and 1.48BPM for three AR/VR headsets. Similarly, the absolute errors of heartbeat rate derivation are improved from 2.77BPM, 3.62BPM, and 3.87BPM to 43.32BPM, 38.90BPM, and 36.78BPM, which demonstrates the effectiveness of our design in mitigating the artifacts of body movements.

7 POTENTIAL DEFENSES

Permissions on AR/VR Motion Sensors. Our attack, reliant on vital sign reconstruction, is viable on prevalent AR/VR programming platforms where access to motion sensor data doesn't necessitate any permission. We demonstrate the potential privacy breaches it could entail, involving sensitive information like gender, identity, and body fat data. To defend *FaceReader*, designers of AR/VR OS platforms should implement a secure and user-friendly permission control interface. This interface would inform users about potential access to malicious motion sensor data. The permission model could involve a run-time prompt, asking for user consent, and any suspicious permission requests should undergo review as part of the AR/VR application review policy.

Privacy-aware Sensor Management. To prevent potential attacks through AR/VR built-in sensors, a potential defense strategy is to develop a privacy-aware sensor management framework. This framework aims to enhance the transparency of sensor utilization on AR/VR headsets. Specifically, the framework could provide real-time statistical assessments of sensor usage, including the duration of sensor use by a specific AR/VR application/website. Based on the design of this sensor management framework, AR/VR users can determine whether the app/website recording sensor data in the background needs to be enforced or not. The framework could also identify suspicious apps or websites continuously recording sensor readings and provide reminders or warnings to AR/VR users.

8 RELATED WORK

Most initial studies of AR/VR privacy leakage focus on active virtual engagement [6, 24, 27], where the adversaries obtain private attributes through users' direct input with personal information (e.g., users' names, gender) in AR/VR scenarios. For example, Maloney *et al.* [24] demonstrate that users often actively reveal their personal

Table 1: Key differences of *FaceReader* from previous works using motion sensors for privacy information derivation.

Works	Leakage Type	No Speech/ Gesture Input	No Users' Data for Training	Longitudinal Tracking	Cross App/ Website Attack
FaceReader	**Vital Sign and Embedded Privacy**	✔	✔	✔	✔
Nair *et al.* [27]	Anthropometric Demographic	✗	✗	✗	✗
Face-Mic [36]	Speech Content Demographic	✗	✔	✗	✗
Wu *et al.* [43]	Typed Characters Passwords	✗	✗	✗	✗
TyPose [37]	Sentences	✗	✔	✗	✗

experiences and information on AR/VR social media platforms. These self-disclosures inevitably raise privacy concerns regarding personal information leaks in AR/VR scenarios. Similarly, Dick *et al.* [6] showcase social engineering attacks that manipulate users into divulging their locations, appearances, and movements. Nair *et al.* [27] design an innocuous-looking yet adversarial 'escape room' game. While the game is running, it collects users' active inputs for solving puzzles (e.g., reading words, pressing virtual buttons, mimicking poses), thereby revealing users' locations and movements and allowing adversaries to infer users' height and demographics (e.g., age and gender). However, all these attacks necessitate considerable skills and effort to manipulate users into revealing private information, making the process complex and time-consuming. Furthermore, the suspicious statements and behaviors of adversaries during virtual interactions may alert users and expose malicious intentions, particularly when adversaries are strangers.

More recently, a few research studies have shown private leakage through unrestricted position and motion sensors on AR/VR devices [36, 37, 43]. For instance, Face-Mic [36] leverages motion sensors on AR/VR headsets to infer users' speech contents and demographic information (e.g., identity, gender). The authors demonstrate that AR/VR motion sensors are capable of capturing facial dynamics induced by human speech. Another example is Wu *et al.* [43], who design a keystroke snooping attack that leverages the position and orientation of AR/VR controllers to infer users' keystrokes, thereby deducing passwords or typed contents. Similarly, Typose [37] explores the feasibility of inferring AR/VR keystrokes from users' head motions. The concept is based on the idea that users' heads move subtly while typing on a virtual keyboard, which can be captured by motion sensors to infer the text input. However, all these attacks are effective only when users interact with the AR/VR systems, such as engaging in speech activities or typing through gestures. In contrast, *FaceReader* presents the first fully passive attack that leverages vital signs naturally produced by the human body, imposing no restrictions on users' activities and thereby remaining unobtrusive. The key differences of *FaceReader* compared to the existing attacks are illustrated in Table 1.

9 CONCLUSION

In this paper, we investigate a stealthy privacy attack called *FaceReader*, which reconstructs fine-grained vital signs based on passive vital-sign-induced facial vibrations collected with unrestricted motion sensors. We showcase three advanced attacks that reveal users' gender, identity, and body fat ratio of victims based on the reconstructed vital signs. Particularly, we design a waveform reconstruction model that utilizes LSTM and a self-attention mechanism to derive high-quality breathing and heartbeat signals. Through adaptive training with a domain discriminator, *FaceReader* can perform the reconstruction without requiring training data or ground truth from the target individual. By leveraging the reconstructed vital sign signals, we design time-frequency features and employ advanced deep-learning models for gender detection and body fat ratio estimation without training data from the victims. Extensive experiments show that *FaceReader* is feasible under various practical attack scenarios and settings (e.g., attack across different users and devices). We further demonstrate practical user re-identification across apps, websites, and longitudinal sessions by designing a CNN-based model. *FaceReader* thus highlights a previously unexplored avenue of privacy leakage through unrestricted AR/VR motion sensors, emphasizing the need for well-directed defense strategies within the AR/VR community.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation Grants CNS2120396, CNS2145389, CNS2120276, CNS2201465, CNS2152669, CCF2000480, CCF2211163, OAC2139358 and IIS2311596.

REFERENCES

[1] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. 2021. Spearphone: A Lightweight Speech Privacy Exploit via Accelerometer-Sensed Reverberations from Smartphone Loudspeakers. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '21)*. Association for Computing Machinery, New York, NY, USA, 288–299. https://doi.org/10.1145/3448300.3468499

[2] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. 2020. Learning-based Practical Smartphone Eavesdropping with Built-in Accelerometer. In *Network and Distributed System Security Symposium*.

[3] Mayo Clinic. 2023. Prescription weight-loss drugs - Study the pros and cons of medicines to treat obesity. https://www.mayoclinic.org/healthy-lifestyle/weight-loss/in-depth/weight-loss-drugs/art-20044832. (2023).

[4] Shifan Dai, Janet E. Fulton, Ronald B. Harrist, Jo Anne Grunbaum, Lyn M. Steffen, and Darwin R. Labarthe. 2009. Blood Lipids in Children: Age-Related Patterns and Association with Body-Fat Indices: Project HeartBeat! *American Journal of Preventive Medicine* 37, 1, Supplement (2009), S56–S64. https://doi.org/10.1016/j.amepre.2009.04.012 Development of Cardiovascular Risk Factors and the Role of Obesity and Related Measures.

[5] Seyedmohsen Dehghanojamahalleh and Mehmet Kaya. 2019. Sex-Related Differences in Photoplethysmography Signals Measured From Finger and Toe. *IEEE J Transl Eng Health Med* 7 (Aug. 2019), 1900607.

[6] Ellysse Dick. 2021. Balancing User Privacy and Innovation in Augmented and Virtual Reality. https://api.semanticscholar.org/CorpusID:232970463

[7] Common Sense Education. 2023. AR and VR Games and Apps for Learning. https://www.commonsense.org/education/lists/ar-and-vr-games-and-apps-for-learning. (2023).

[8] Mona A. Eissa, Shifan Dai, Nicole L. Mihalopoulos, R. Sue Day, Ronald B. Harrist, and Darwin R. Labarthe. 2009. Trajectories of Fat Mass Index, Fat Free–Mass Index, and Waist Circumference in Children: Project HeartBeat! *American Journal of Preventive Medicine* 37, 1, Supplement (2009), S34–S39. https://doi.org/10.1016/j.amepre.2009.04.005 Development of Cardiovascular Risk Factors and the Role of Obesity and Related Measures.

[9] Emre Ertin, Nithin Sugavanam, August Holtyn, Kenzie Preston, Jeremiah Bertz, Lisa Marsch, Bethany McLeman, Dikla Shmueli Blumberg, Julia Collins, Jacqueline King, Jennifer Mccormack, and Udi Ghitza. 2021. An Examination of the Feasibility of Detecting Cocaine Use Using Smartwatches. *Frontiers in Psychiatry* 12 (06 2021), 674691. https://doi.org/10.3389/fpsyt.2021.674691

[10] exyte. 2019. Technologies behind immersive VR: positional tracking and VR-accessories. (2019). https://arvrjourney.com/technologies-behind-immersive-vr-positional-tracking-and-vr-accessories-part-1-f38f9a7f1a02

[11] Bracket Foundation. 2022. Gaming and the Metaverse: The Alarming Rise of Online Sexual Exploitation and Abuse of Children Within the NEw Digital Frontier. https://www.weprotect.org/wp-content/uploads/Gaming_and_the_Metaverse_Report_final.pdf

[12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2017.

Domain-Adversarial Training of Neural Networks. Springer International Publishing, Cham, 189–209. https://doi.org/10.1007/978-3-319-58347-1_10
[13] Gonzalo Munilla Garrido, Vivek Nair, and Dawn Song. 2023. SoK: Data Privacy in Virtual Reality. (2023). arXiv:cs.HC/2301.05940
[14] Google. 2023. Android Developers: Android Mobile App Developer Tools. (2023). https://developer.android.com/
[15] USA GOV. 2023. Privacy and security policies. https://www.usa.gov/privacy. (2023).
[16] Alexander Hermans, Lucas Beyer, and Bastian Leibe. 2017. In Defense of the Triplet Loss for Person Re-Identification. (2017). arXiv:cs.CV/1703.07737
[17] HP. 2023. HP Reverb G2 Omnicept Edition. https://www.hp.com/us-en/vr/reverb-g2-vr-headset-omnicept-edition.html. (2023).
[18] Pengfei Hu, Hui Zhuang, Panneer Selvam Santhalingam, Riccardo Spolaor, Parth H. Pathak, Guoming Zhang, and Xiuzhen Cheng. 2022. AccEar: Accelerometer Acoustic Eavesdropping with Unconstrained Vocabulary. In *43rd IEEE Symposium on Security and Privacy, SP 2022, San Francisco, CA, USA, May 22-26, 2022*. IEEE, 1757–1773. https://doi.org/10.1109/SP46214.2022.9833716
[19] Jingyu Hua, Zhenyu Shen, and Sheng Zhong. 2017. We Can Track You if You Take the Metro: Tracking Metro Riders Using Accelerometers on Smartphones. *IEEE Transactions on Information Forensics and Security* 12, 2 (2017), 286–297. https://doi.org/10.1109/TIFS.2016.2611489
[20] Yulong Huang, Yonggang Zhang, and Jonathon A. Chambers. 2019. A Novel Kullback–Leibler Divergence Minimization-Based Adaptive Student's t-Filter. *IEEE Transactions on Signal Processing* 67, 20 (2019), 5417–5432. https://doi.org/10.1109/TSP.2019.2939079
[21] James M. Joyce. 2011. *Kullback-Leibler Divergence*. Springer Berlin Heidelberg, Berlin, Heidelberg, 720–722. https://doi.org/10.1007/978-3-642-04898-2_327
[22] KasperSky. 2023. What are the Security and Privacy Risks of VR and AR. https://usa.kaspersky.com/resource-center/threats/security-and-privacy-risks-of-ar-and-vr. (2023).
[23] Yang Liu and Zhenjiang Li. 2018. aLeak: Privacy Leakage through Context - Free Wearable Side-Channel. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. 1232–1240. https://doi.org/10.1109/INFOCOM.2018.8485958
[24] Divine Maloney, Samaneh Zamanifard, and Guo Freeman. 2020. Anonymity vs. Familiarity: Self-Disclosure and Privacy in Social Virtual Reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology (VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 25, 9 pages. https://doi.org/10.1145/3385956.3418967
[25] Meta. 2021. Oculus Mobile SDK. (2021). https://developer.oculus.com/downloads/package/oculus-mobile-sdk/
[26] Meta. 2023. Horizon Worlds, Virtual Reality Worlds and Communities - Meta. (2023). https://www.meta.com/horizon-worlds/
[27] Vivek Nair, Gonzalo Munilla Garrido, and Dawn Song. 2023. Exploring the Unprecedented Privacy Risks of the Metaverse. (2023). arXiv:cs.CR/2207.13176
[28] Omkar R. Patil, Wei Wang, Yang Gao, Wenyao Xu, and Zhanpeng Jin. 2018. A Non-Contact PPG Biometric System Based on Deep Neural Network. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 1–7. https://doi.org/10.1109/BTAS.2018.8698552
[29] PROPUBLICA. 2023. Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates. (2023). https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates
[30] Rebecca M. Puhl and Chelsea A. Heuer. 2010. Obesity Stigma: Important Considerations for Public Health. *American Journal of Public Health* 100, 6 (June 2010), 1019–1028. https://doi.org/10.2105/ajph.2009.159491
[31] George R. Doddington R. Gary Leonard. 1993. TIDIGITS LDC93S10. https://catalog.ldc.upenn.edu/LDC93S10. (1993).
[32] RENPHO. 2023. Smart Body Weight Scales. https://renpho.com/collections/renpho-scales. (2023).
[33] sensoriumgalaxy. 2023. The Sensorium Galaxy Metaverse - Out-of-This-World Experiences. (2023). https://sensoriumgalaxy.com/
[34] Theodoros N. Sergentanis, Sofia D. Bampalitsa, Paraskevi Theofilou, Eleni Panagouli, Elpis Vlachopapadopoulou, Stefanos Michalacos, Alexandros Gryparis, Loretta Thomaidis, Theodora Psaltopoulou, Maria Tsolia, Flora Bacopoulou, and Artemis Tsitsika. 2021. Cyberbullying and Obesity in Adolescents: Prevalence and Associations in Seven European Countries of the EU NET ADB Survey. *Children* 8, 3 (2021). https://doi.org/10.3390/children8030235
[35] Julie Setele. 2022. Matthew Crain. Profit over Privacy: How Surveillance Advertising Conquered the Internet. Minneapolis, MN: University of Minnesota Press, 2021. 216p. Paperback, $25.00 (ISBN: 978-1-5179-0505-7). *College Research Libraries* 83, 4 (2022), 702. https://doi.org/10.5860/crl.83.4.702
[36] 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. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 478–490. https://doi.org/10.1145/3447993.3483272
[37] Carter Slocum, Yicheng Zhang, Nael Abu-Ghazaleh, and Jiasi Chen. 2023. Going through the motions: AR/VR keylogging from user head motions. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 159–174. https://www.usenix.org/conference/usenixsecurity23/presentation/slocum
[38] Valve Software. 2018. OpenVR SDK. (2018). https://github.com/ValveSoftware/openvr
[39] Statista. 2023. AR/VR - Worldwide Statista Market Forecast. https://www.statista.com/outlook/amo/ar-vr/worldwide. (2023).
[40] TECHVIZ. 2023. 4 use cases for virtual reality in the military and defense industry. https://blog.techviz.net/4-use-cases-for-virtual-reality-in-the-military-and-defense-industry. (2023).
[41] He Wang, Ted Tsung-Te Lai, and Romit Roy Choudhury. 2015. MoLe: Motion Leaks through Smartwatch Sensors. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15)*. Association for Computing Machinery, New York, NY, USA, 155–166. https://doi.org/10.1145/2789168.2790121
[42] WebXR. 2023. WebXR Device API. (2023). https://immersiveweb.dev/
[43] Yi Wu, Cong Shi, Tianfang Zhang, Payton Walker, Jian Liu, Nitesh Saxena, and Yingying Chen. 2023. Privacy Leakage via Unrestricted Motion-Position Sensors in the Age of Virtual Reality: A Study of Snooping Typed Input on Virtual Keyboards. In *2023 IEEE Symposium on Security and Privacy (SP)*. 3382–3398. https://doi.org/10.1109/SP46215.2023.10179301
[44] XRToday. 2023. Five Augmented Reality Apps Revolutionizing Healthcare. https://www.xrtoday.com/augmented-reality/5-augmented-reality-apps-revolutionizing-healthcare/. (2023).
[45] Chenhan Xu, Huining Li, Zhengxiong Li, Hanbin Zhang, Aditya Singh Rathore, Xingyu Chen, Kun Wang, Ming-chun Huang, and Wenyao Xu. 2021. CardiacWave: A MmWave-Based Scheme of Non-Contact and High-Definition Heart Activity Computing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 135 (sep 2021), 26 pages. https://doi.org/10.1145/3478127
[46] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo. 2022. Can We Obtain Fine-grained Heartbeat Waveform via Contact-free RF-sensing?. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*. 1759–1768. https://doi.org/10.1109/INFOCOM48880.2022.9796905