

# Harnessing Vital Sign Vibration Harmonics for Effortless and Inbuilt XR User Authentication

Tianfang Zhang  
Rutgers University  
New Brunswick, New Jersey, USA  
tz203@scarletmail.rutgers.edu

Qiufan Ji  
New Jersey Institute of Technology  
Newark, New Jersey, USA  
qj39@njit.edu

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

Zhengkun Ye  
Temple University  
Philadelphia, Pennsylvania, USA  
zhengkun.ye@temple.edu

Ahmed Tanvir Mahdad  
Texas A&M University  
College Station, Texas, USA  
mahdad@tamu.edu

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

Yan Wang  
Temple University  
Philadelphia, Pennsylvania, USA  
y.wang@temple.edu

Nitesh Saxena  
Texas A&M University  
College Station, Texas, USA  
nsaxena@tamu.edu

Yingying Chen\*  
Rutgers University  
New Brunswick, New Jersey, USA  
yingche@scarletmail.rutgers.edu

## Abstract

Extended Reality (XR) headsets are increasingly serving as repositories for substantial volumes of sensitive data and gateways to web applications. This transition highlights the need for convenient and secure user authentication solutions. Traditional password/PIN-based schemes are ill-suited to the XR's gesture- and voice-based interfaces and are prone to shoulder-surfing attacks. Some recent XR systems incorporate two-factor authentication, but it requires additional operations on a second device (e.g., a smartphone or wearable). In this work, we introduce the first effortless and inbuilt XR user authentication system by leveraging the harmonics of vibrations excited by users' vital signs. The system is transparent to users (no efforts during enrollment and authentication) and requires no additional hardware. The key idea is that vital signs (i.e., breathing and heart beating) naturally generate low-frequency mechanical vibrations, causing human skull to vibrate and produces harmonic signals. When the harmonics pass the human head, they carry rich biometrics associated with the wearer's skull structure and soft tissues, which can be captured by the XR motion sensors. Instead of directly utilizing the vibrations, we extract more reliable biometrics from the ratios among different harmonic frequencies, which capture wearers' unique head and facial attenuation properties and are non-volatile when the periodicity and amplitude of vital signs fluctuate. We further design an adaptive filter to mitigate the body motion distortions in common XR interactions. By adopting advanced deep learning models with the attention mechanism, our system realizes effective and robust authentication across XR

scenarios. Evaluations across 10 months, with 52 users and two popular XR headsets, show that our system can accurately authenticate users with over 95% true positive rates and rejects unauthorized users with over 98% true negative rates under various XR scenarios, with biometrics remaining consistent over long-term periods.

## CCS Concepts

• Security and privacy → Authentication.

## Keywords

XR headsets, Authentication, Vital sign harmonics

### ACM Reference Format:

Tianfang Zhang, Qiufan Ji, Md Mojibur Rahman, Redoy Akanda, Zhengkun Ye, Ahmed Tanvir Mahdad, Cong Shi, Yan Wang, Nitesh Saxena, and Yingying Chen. 2025. Harnessing Vital Sign Vibration Harmonics for Effortless and Inbuilt XR User Authentication. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25)*, October 13–17, 2025, Taipei, Taiwan. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3719027.3765060>

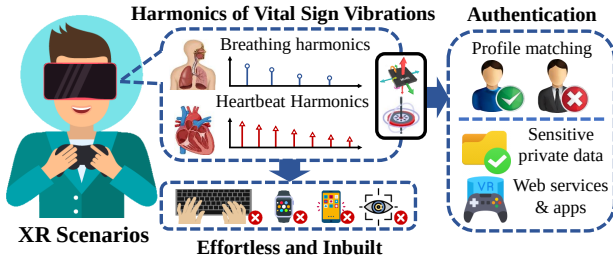
## 1 Introduction

Head-mounted Extended Reality (XR) devices have rapidly gained global attention, driven by advancements in 3D spatial displays and motion tracking technologies. These devices allow users to interact with programs and digital media through intuitive inputs such as gestures and voice in augmented (AR), virtual (VR), and mixed realities (MR) [3, 17, 18]. The adoption of XR headsets grows rapidly at an annual rate of 11.5% and will result in more than 34 million users in 2026 [1]. The rapid surge of XR devices is driving a transition from traditional computers, limited by 2D screens or mice/keyboards, to spatial computing using XR headsets [7] and reforming many critical sectors, such as remote workspace/collaboration [6], manufacturing [2], education [9], entertainment [8], and healthcare [19]. As the XR headsets are gaining worldwide adoption, a substantial amount of sensitive data (e.g., personal and work accounts, confidential documents, financial and health records) is being transferred

\*Yingying Chen is the corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.



**Figure 1: Illustration: Our system harnesses harmonics of vital sign vibrations to realize effortless (does not require any explicit user actions) and inbuilt (uses motion sensors available on all XR headsets) XR user authentication.**

from traditional computing devices to XR headsets. They are also increasingly serving as gateways to many web services (e.g., search engines, online shopping) and applications (e.g., editors, games, virtual assistants). These transitions highlight the urgent need for convenient and secure user authentication for XR headsets.

However, the development of user authentication for XR systems still remains in its infancy. Existing password/PIN-based methods directly migrated from traditional computing devices do not align well with XR’s unique input interfaces. Utilizing gestures or voices to enter passwords or PINs in XR systems is usually cumbersome as XR devices normally do not include a physical keyboard. These methods are also prone to the observation and shoulder-surfing attacks [21, 48, 51]. To address these issues, behavioral/physiological biometrics [38, 44, 52, 53, 63] and two-factor authentication (2FA) [10, 11, 13] have been investigated. Nevertheless, these methods present two main limitations: (i) *Lack of Seamlessness*: Some XR headsets have adopted a QR code [10] or virtual button [11] on a smartphone/wearable as a second factor for authentication. However, they usually require operations on the second device, interfering with XR’s immersive experience. A few high-end XR headsets (e.g., Apple Vision Pro) employ infrared cameras for iris authentication, but such headsets are prohibitively expensive (e.g., over \$3000). More affordable headsets, such as Meta Quest 2/3 and Valve Index, typically lack these dedicated sensors. (ii) *Lack of Biometrics Robustness*. Recent studies leverage behavioral biometrics from the user’s hand [52, 53], head [38, 63], and eye motions [44] for authentication. This trend brings convenience as gestures are commonly used during XR interactions. However, these mechanisms experience reduced performance due to intra-user variability, meaning that a user’s behavior may shift with different emotional states, physical conditions, XR contexts, etc.

In this paper, we set forth to address the limitations by designing a seamless and robust user authentication system for XR headsets. We find that vital signs (i.e., breathing and heartbeating) naturally generate low-frequency mechanical vibrations propagating through the human head. These vibrations cause the skull structure to vibrate, producing the signal patterns at the natural frequencies (i.e., multiples of the breathing/heartbeat frequency), which we refer to as *vital sign harmonics*. These vibration harmonics reflect the unique skull structural properties of the user, such as geometry and density of cranial bones, facial bones, and ossicles. In addition, the vibration signals propagate through the soft tissues of the face (e.g., skin, muscle, fat), undergoing a unique attenuation pattern shaped

by the face’s bio-mechanical properties. As the headset is mounted on the user’s head, it can pick up the vital sign harmonics, thereby passively encoding the skull and facial tissue properties into the motion sensor readings. We show that robust biometrics can be extracted from the vital sign harmonics even when the breathing and heartbeat rates change. The robustness is rooted in the consistent properties of skull structure and facial tissues, enabling non-volatile biometric extraction. More importantly, the vibration harmonics carry head properties confined to the human body. The internal propagation mechanism makes our system resilient to spoofing attacks relying on biometric theft via video/voice recordings.

The core idea of our system is to derive robust biometric features from vital sign harmonics to realize user authentication, as illustrated in Figure 1. The user enrollment and authentication process of our system are transparent. As built-in XR motion sensors are always on to track head motions, our system can extract biometrics by analyzing streaming sensor readings for either user profile construction or authentication, without requiring additional devices or sensors. Compared with traditional password/PIN-based user authentication methods, our system removes the need for explicit user inputs, such as entering passwords, interacting with secondary devices, and performing gestures, which could interrupt or negatively impact the user’s XR experience. While a recent study [69] shows the feasibility of using facial vibrations induced by vital signs for user re-identification, it did not account for the inherent physiological variability in vital signs [49, 55], such as fluctuations in periodicity, amplitude, and temporal patterns. For instance, a user’s heart rate and pulse volume typically elevate during intense gaming compared to more passive activities, such as watching 3D videos. Similarly, immersive XR experiences may cause shorter respiration cycles if users become excited or anxious. These vital sign variations may cause a biometric mismatch between the user’s profile constructed during the enrollment and the authentication phase with new vibration data. Instead of directly utilizing facial vibrations, we extract biometric features from various harmonic components that reflect the stable biomechanical properties of skull structures and facial tissues. Particularly, we leverage ratios between harmonic components of the vibrations. These ratio-based features capture unique skull structure properties [25, 31], such as the geometry and density of the skull. Furthermore, the unique attenuation properties of soft tissues of the face are also reflected in these features. Both types of properties are non-volatile under the variations of vital signs, thereby allowing robust user authentication.

To realize such a system, we need to address several challenges. (1) *Distortions Caused by Body Motions*. In XR contexts, the user interacts with virtual objects via various body motions (e.g., head and hand gestures). The motion artifact can affect signal patterns of a wide frequency range, even overshadowing the vital sign harmonics. (2) *Unknown Biometrics Related to Harmonic Ratio*. Biometric features based on harmonic attenuation ratio have not been studied in prior works. It is necessary to study the correlations among different harmonic components, and extract features invariant to vital sign variations. (3) *Biometric Robustness Across XR Contexts*. Vital sign harmonics and their ratios are inherently sequential signals, with small-scale temporal variations. To realize robust user authentication, our system should derive reliable representations capturing inherent skull structure and tissue properties.

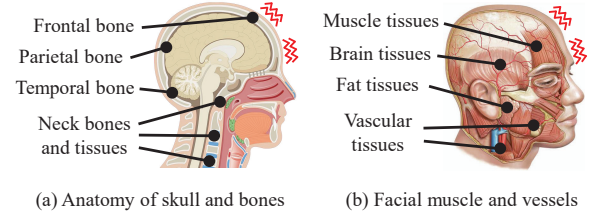
To mitigate the distortions in vital sign harmonics, we develop an adaptive filtering scheme that dynamically cancels the motion artifacts and extracts the denoised harmonic signals. We further develop a harmonic augmentation scheme that simulates the distortions caused by variations in vital sign magnitude and frequency to enrich user profiles for robust authentication. Based on the consistent harmonic ratio under different physiological conditions, we develop a series of attenuation features, such as Harmonic Amplitude Ratio and Harmonic Energy Ratio, to depict users' unique skull and facial soft tissues. Additionally, we derive structural features, including cepstral and wavelet features, to characterize vital sign harmonics from the spatial domain. Together, these features and augmentation scheme enable robust user profile construction that accommodates various XR scenarios and long-term usage. We further design a transformer-based model with self-attention mechanisms to derive user representations. A user will be authenticated if the representations match those in the claimed identity profile. Our main contributions can be summarized as follows:

- We propose the first effortless and inbuilt user authentication system for commodity XR headsets harnessing unique head biometrics derived from vital sign harmonics. It is also the first work that demonstrates distinctive and robust biometrics can be extracted from vibration harmonics, without requiring any active user interactions or extra hardware/sensors.
- We design a series of harmonic features based on the ratios between different harmonic components of vital sign vibrations. These harmonic features characterize users' unique skull structures and properties of soft tissues on human faces, which are non-volatile under the variations of vital signs.
- We develop a harmonic augmentation approach based on noise injection and frequency shifting to simulate the distortions caused by vital sign variations and enrich user biometric profiles. For reliable authentication, we develop a transformer-based encoder to derive robust feature representations that characterize the skull and facial tissue properties of the wearer.
- We validate our system by conducting extensive experiments on Meta Quest and HTC Vive Pro Eye across 10 months with more than 52 users aged 18 to 45. The results show that our system achieves true positive rates of over 95% under various XR scenarios and long-term authentication (e.g., over 60 days). Moreover, it effectively rejects adversaries with true negative rates of more than 98% under practical attack scenarios.

## 2 Preliminary Studies

This work aims to explore effective and robust user authentication by analyzing vital sign harmonics that characterize human skull and facial soft tissues of the wearer. As illustrated in Figure 2, the human head consists of skull bones, such as the frontal, parietal, and temporal bones. When the mechanical vibrations induced by vital signs propagate through the head, the human skull will act as a resonant structure, and produce distinctive harmonic patterns. According to prior studies [37, 57, 62], the relationship among the amplitudes of fundamental frequencies and harmonics of vital sign vibrations on the human skull can be formulated as:

$$A(i \cdot f_0) \propto A(f_0) \cdot \frac{E}{\rho \cdot \eta \cdot i^k}, \quad k > 1, \quad (1)$$



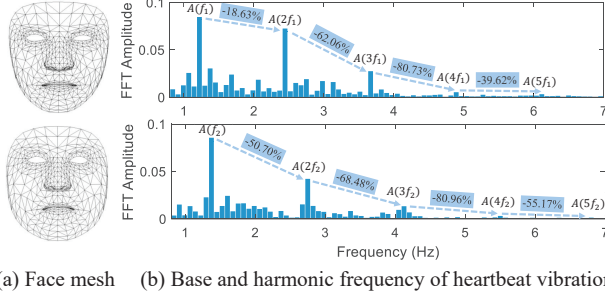
**Figure 2: Kinetics and related structures of the minute vibrations induced by human respiration and heartbeat.**

where  $A(\cdot)$  denotes the frequency amplitude of vital sign harmonics on the human skull and  $f_0$  represents the fundamental frequency of vital sign vibrations.  $E$ ,  $\rho$ , and  $\eta$  refer to the elastic modulus, density, and viscosity.  $i$  is the harmonic level and  $k$  is an empirical constant, which indicates the exponential reduction of harmonic amplitude on higher frequencies. The bio-mechanical properties of human skull and head bones [20, 22, 58] are summarized in Table 1. With extremely high elastic modulus and low viscosity, the human skull tends to amplify the harmonics after receiving the vital sign vibrations. Moreover, users' unique skull structures and properties are encoded in the distinct harmonic amplitudes after amplification, enabling biometric extraction from vital sign harmonics.

After being manipulated by the human skull, the vital sign harmonics travel through the soft tissues of human faces, manifest distinct attenuation patterns. With different bio-mechanical properties in Table 1, these tissues create a complex medium for propagating vibrations. In particular, soft tissues, such as brain, muscle, and fat, have substantially low elastic modulus and high viscosity compared to the human skull. According to Equation 1, the amplitude of  $i$ -th vital sign harmonic  $A(i \cdot f_0)$  will experience unique attenuation effects while passing the different facial tissues given their distinct bio-mechanical properties. Inspired by these findings, we explore how skull structures and tissue properties influence the harmonic generation and attenuation signatures of individual users.

**Harmonic Properties of Vital Signs.** To investigate the impact of skull structures and facial tissues on the vital sign harmonic patterns and attenuation effect, we conduct a preliminary experiment using Meta Quest. Specifically, we recruit two participants (U1 and U2) with different facial structures and distributions of tissues. During the experiment, both participants wear the headset, sit still and watch an immersive XR spacewalk video. The sampling rate of the headset's accelerometer is set to 1000Hz. The face meshes and frequency distributions of heartbeat harmonics from both participants are shown in Figure 3. In particular, U1, whose face has a higher proportion of bones and limited fat tissues, manifests strong harmonic amplitudes and mild attenuation effects (e.g., a reduction of 18.63% from the fundamental amplitude  $A(f_1)$  to the first harmonic amplitude  $A(2f_1)$ ). In contrast, U2, whose face contains more fat tissue, shows weaker harmonic amplitudes and significant harmonic attenuation (e.g., a reduction of 50.70% from the fundamental amplitude  $A(f_2)$  to the first harmonic amplitude  $A(2f_2)$ ). Moreover, compared to U1, U2's face indicates more prominent attenuation at higher frequencies, which aligns with the analysis from Equation 1. We further analyze the multi-dimensional motion sensor readings and observe that facial width and skull thickness influence the harmonic attenuation patterns along the y- (parallel to the human faces) and z-axes (perpendicular to the human faces).





**Figure 3: Face structures of two users (U1 and U2) and their harmonic attenuation patterns of heartbeat vibrations.**

**Table 1: Bio-mechanical characteristics (e.g., density, elasticity, viscosity) of representative facial tissues and structures.**

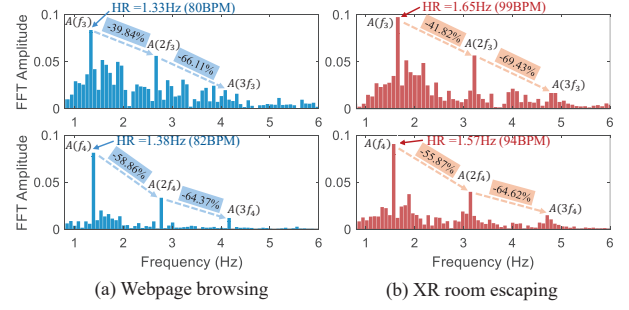
| Facial tissues | Density ( $kg/m^3$ )    | Elasticity ( $kPa$ )                     | Viscous Ratio (%) |
|----------------|-------------------------|------------------------------------------|-------------------|
| human skull    | $\sim 1.92 \times 10^3$ | $1.00 \times 10^6 \sim 7.00 \times 10^6$ | $\sim 15\%$       |
| head bones     | $\sim 1.90 \times 10^3$ | $1.00 \times 10^5 \sim 5.00 \times 10^5$ | $\sim 20\%$       |
| brain tissue   | $\sim 1.04 \times 10^3$ | $0.1 \sim 100$                           | $40\% \sim 50\%$  |
| facial skin    | $\sim 1.10 \times 10^3$ | $1.95 \times 10^4 \sim 8.71 \times 10^4$ | $30\% \sim 60\%$  |
| facial muscle  | $\sim 1.06 \times 10^3$ | $6.40 \times 10^2 \sim 1.58 \times 10^3$ | $18\% \sim 22\%$  |
| fat tissue     | $\sim 0.90 \times 10^3$ | $5.3 \sim 18.1$                          | $45\% \sim 60\%$  |

Specifically, users with wider faces and thicker skull structures exhibit stronger attenuation effects from the base frequency to the first-order harmonic. These findings validate that vibration harmonics can capture the anatomical differences of human faces from spatial perspectives and attenuation patterns of vital sign vibrations can be leveraged to differentiate users.

**Consistency of Harmonic Properties.** The patterns of vital sign excitations, such as periodicity, amplitude, and morphology, naturally change with the users' physiological status. Different XR scenarios, such as glancing over and being attentive to XR contents, could influence the vital sign patterns, and alter the energy and frequency distribution of vital sign harmonics. To study the impacts of vital sign variations on harmonic properties, we collect the data from two participants (U3 and U4) under two XR scenarios: 1) browsing an XR webpage, and 2) playing an XR room escaping game. Compared to webpage browsing, the immersive room escape game leads to a prominent increase in the user's heart rate. The harmonic distributions for each scenario are shown in Figure 4. Despite physiological changes, the attenuation ratio across the amplitude of fundamental frequency, first and second harmonic remains stable for the same user's facial properties. Nevertheless, the overall energy is generally higher during the room escape game due to the increased pulse rates. These differences in harmonic frequency and energy distribution pose challenges for robust biometric derivation. To realize reliable user authentication, we develop a set of features to characterize human skull and facial attenuation properties of individual users, which will be introduced in Section 5.2.

### 3 Threat Model

The adversary understands the basic workflow of our system and attempts to bypass it to access the private data (e.g., photos, medical/financial records), web services, or applications linked to the user's XR device. We consider the following attacks with different prior knowledge of the legitimate user's vital sign harmonics.



**Figure 4: Harmonic attenuation of heartbeat vibrations from two users (U3 and U4) under different XR scenarios.**

**Scenario I: Attack with Zero Prior Knowledge.** In this attack, the adversary does not have any prior knowledge of the harmonic attenuation signatures of legitimate users. During the attack, the adversary wears the user's XR headset, and uses his own vital signs to generate similar vibration patterns for bypassing our system. The adversary may also recruit other people for this attack, hoping to bypass our authentication system.

**Scenario II: Attack with Prior Knowledge of Vital Signs.** The adversary may have prior knowledge of the legitimate users' respiration rates/patterns (e.g., captured via a respiration monitoring belt), heartbeat rates/patterns (e.g., captured via a PPG sensor), demographic information (e.g., gender and age), and body measurements (e.g., height, weight, and fat ratio), which could be leaked from their health records. The adversary can recruit people with similar demographic and body measurements to those of legitimate users and instruct them to imitate the breathing rates (e.g., using a metronome). In terms of heartbeat mimicking, as heartbeats are uncontrollable, the adversary may choose people with similar demographic and body measurements and then picks those with resting heart rates similar to the legitimate users to perform the attack. By doing these, the adversary expects to replicate the attenuation patterns of vital sign harmonics from legitimate users, and utilize them to bypass the authentication of our system.

Note that we do not consider OS/firmware attacks against our system, where an adversary digitally feeds pre-collected harmonic attenuation signatures for spoofing. All biometric authentication systems could face security breaches if the OS/firmware is compromised. They are beyond the scope of our work studying biometric authentication based on human skull and facial tissue properties.

## 4 System Design

### 4.1 Challenges

**Significant Distortions from Body Motions.** In common XR scenarios, users could frequently interact with the XR system via different types of gestures (e.g., moving controllers, rotating head, grabbing virtual objects). Such movements could distort the minute vibrations associated with human vital signs and their harmonic patterns. To achieve reliable authentication under different XR scenarios, we should develop an effective scheme to mitigate the impacts of distortions caused by body motions.

**Deriving Biometrics Related to Head Properties.** The relationship between the attenuation ratio of vital sign harmonics and human skull/facial tissue properties has not been studied in

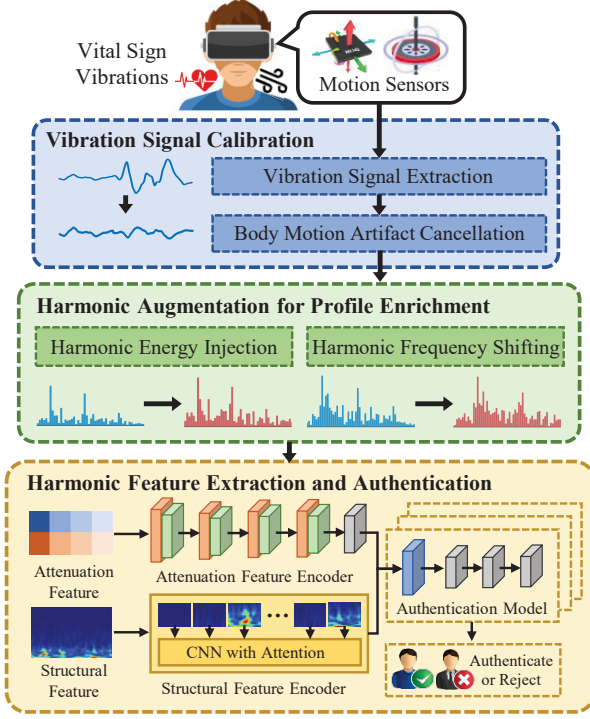


Figure 5: Overview of the authentication system.

previous studies. It is essential to extract biometric features that remain consistent under the variations of vital sign excitations, while carrying distinctive head characteristics (e.g., structure of human skull, attenuation of soft tissues) for user authentication.

**Biometric Robustness Across XR Contexts.** Vital sign harmonics and the attenuation ratios among different harmonic components are embedded into sequential motion sensor readings. Thus, the biometrics extracted from the XR sensor readings inherit some temporal variations of vital signs, even after biometric feature extraction. The system should derive more reliable representations that maintain a high consistency.

## 4.2 System Overview

The basic idea of our system is to derive unique human skull structures and facial soft tissue properties from vital sign harmonics. As illustrated in Figure 5, our system has three key components, including *Vibration Signal Calibration*, *Harmonic Feature Extraction*, and *User Authentication based on Attention*.

**Vital Sign Signal Calibration.** To derive clean harmonic patterns without body motion artifacts, we develop an adaptive filtering approach. This approach dynamically removes motion distortions and reconstructs clean patterns of vital sign harmonics. To separate the breathing and heartbeat patterns, we apply two band-pass filters with different cut-off frequencies to separate the harmonics. The system then calculates the 3D velocity and displacement of two types of harmonics to depict their geometric patterns. The breathing and heartbeat harmonics are then split into segments with a sliding window for further processing.

**Harmonic Feature Extraction.** We first propose a harmonic augmentation scheme based on energy injection and frequency

shifting to simulate the energy fluctuations and frequency changes caused by vital sign variations under different XR scenarios. Inspired by the consistent attenuation ratio across different levels of harmonics, we develop a series of attenuation features, such as Harmonic Amplitude Ratio and Harmonic Energy Ratio, to characterize the unique bio-mechanical properties of users' facial soft tissues. During the internal propagation of vital sign harmonics, the structure of human skull and face also affect the patterns of harmonic attenuation. To characterize the unique internal structures of users' skull and facial tissues, we extract structural features, including cepstrogram features and wavelet features. By incorporating these harmonic-based features and the augmentation scheme, we enhance the robustness of user profile construction under various physiological conditions caused by different XR contexts.

**User Authentication Based on Attention.** With attenuation and structural features extracted from vital sign harmonics, distinct deep-learning-based encoders are then designed to derive unique user embeddings from these features. For attenuation features, we develop a Long Short-Term Memory (LSTM) network with the self-attention mechanism to extract user embeddings. For structural features, we propose a feature encoder based on the transformer architecture, which splits the 2D patterns into patches with temporal position encoding. After generating user representations by combining both feature embeddings, a binary classifier is built for each legitimate user to distinguish him/her from others. Moreover, we accommodate multi-user enrollment (e.g., family use of XR devices) by creating individual profiles for all legitimate users, with each represented by a unique binary classifier.

## 5 Vital Sign Biometric Extraction

### 5.1 Vital Sign Harmonic Calibration

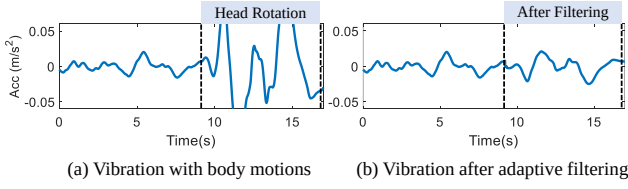
In practical XR scenarios, the motions that users perform to interact with the headsets (e.g., hand gestures, head rotations) could distort vital sign harmonics across a wide range of frequencies. Given the strong instantaneous acceleration of these motions, the vital sign harmonics could be completely overshadowed by these motions and cannot be effectively recovered via simple bandpass filtering. To mitigate the motion artifacts, we develop a vital sign harmonic calibration scheme based on adaptive filtering.

**Body Motion Detection via Short-Time Energy.** Our approach calculates the Short-Time Energy (STE) on the time series of the 3-axis motion sensor readings using a sliding window with a pre-defined size to locate the temporary regions of body motions. With all axes of the accelerometer readings  $x(t)$  captured from the XR headsets as inputs, the STE computation can be described as:

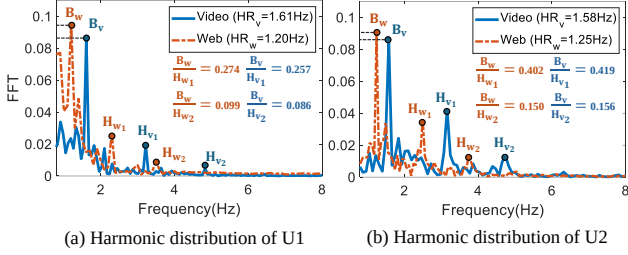
$$STE(x(t)) = \sum_{t=0}^{t+N} x(t)^2, t \in \{0, 1, 2, \dots, L - N\}, \quad (2)$$

where  $L$  is the length of accelerometer reading  $x(t)$ .  $N$  denotes the size of the pre-defined sliding window and we set  $N = 1000$  (e.g., 1-second data with the sampling rate of 1000Hz). We then apply an empirical threshold  $\tau = 0.1$ , which is pre-defined from extensive data collection under various scenarios, to determine whether body motions exist in the current sliding window.

**Design of Adaptive Filtering Scheme.** Then, an adaptive filter is applied to the detected regions to mitigate the motion artifacts



**Figure 6: Examples of breathing vibrations before and after applying our proposed adaptive filtering method.**



**Figure 7: Harmonic distributions on heartbeat vibrations from two different users under different XR scenarios.**

and reconstruct clean vital sign harmonics. The adaptive filtering scheme can be formulated as an optimization problem. Specifically, we define an adaptive filter via a weight vector  $w$  with the same length  $N$  of the sliding window. For signal segments with body motions, we locate the adjacent segment with the same length that is not affected by body motions (e.g., STE lower than the threshold  $\tau$ ). The segment then works as the reference signal to optimize the weight  $w$  until the segments with body motions have similar patterns to the reference signal, which can be formulated as:

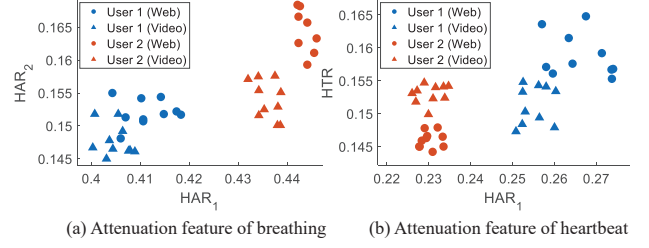
$$\arg \min_w \sum_{t=0}^{L-N} E(t) = D_{KL}(r(t) \parallel \hat{x}(t)) = r(t) \cdot \log \frac{r(t)}{\hat{x}(t)}, \quad (3)$$

s.t.  $w(t) = \alpha \cdot w(t) + \mu \cdot L(t) \cdot x(t)$ ,  $\hat{x}(t) = w(t) \cdot x(t)$ ,

where  $x(\cdot)$ ,  $\hat{x}(\cdot)$ , and  $r(\cdot)$  are the motion sensor readings before and after motion artifact mitigation, and the reference signal.  $\alpha$  and  $\mu$  refer to the leakage parameter for initiating the weight  $w$  and the step size for optimization. To optimize the weights of adaptive filter, we apply Kullback-Leibler Divergence [33] as the loss function given its reliable quantification of similarity between two signal patterns. Examples in Figure 6 validate that the motion artifacts can be effectively mitigated via adaptive filtering. We then separate the denoised breathing and heartbeat harmonics using two Butterworth band-pass filters with the cut-off frequencies of normal breathing (0.1Hz ~ 0.5Hz) and heartbeat (0.8Hz ~ 3.0Hz). The filtered vibration harmonics are split into segments with the same length (e.g., 3 seconds) to derive user biometrics.

## 5.2 Harmonic Feature Extraction

**Attenuation Feature Extraction.** During the propagation of vital sign vibrations, users' facial soft tissues manifest distinct attenuation patterns on vital sign harmonics. Given the consistent bio-mechanical properties of users' facial tissues, the attenuation ratio across different levels of harmonics remains stable, even with different fundamental frequencies (i.e., breathing and heartbeat frequencies). To validate this finding, we collect the frequency distribution of heartbeat harmonics from two users (e.g., U1 and U2)



**Figure 8: Distribution of the attenuation features from two users under different XR scenarios.**

under two XR scenarios (e.g., browsing an XR webpage and watching a spacewalk video), as shown in Figure 7. In particular, U1's heart rate while browsing the webpage is 1.20Hz (72BPM), with a ratio of 0.274/0.099 between the amplitude of fundamental frequency and first/second harmonics. Although the heart rate increases to 1.61Hz (97BPM) while watching the video, the ratio between the amplitude of fundamental frequency and first/second harmonics remains 0.257/0.086 without significant changes. Inspired by the consistency of attenuation ratio among different harmonics, we derive the Harmonic Amplitude Ratio (HAR) and Harmonic Energy Ratio (HER) from harmonic signals, which can be formulated as:

$$\text{HAR}_i(x(t)) = \frac{A_{h_i}(x(t))}{A_b(x(t))}, \quad \text{HER}_i(x(t)) = \frac{P_{h_i}(x(t))}{P_b(x(t))}, \quad (4)$$

where  $A_{h_i}(\cdot)$  and  $A_b(\cdot)$  denote the amplitude of Short-Term Fourier Transform (STFT) from the  $i$ -th harmonic and fundamental frequency.  $P_{h_i}(\cdot)$  and  $P_b(\cdot)$  refer to the power of  $i$ -th harmonic and fundamental frequency. Specifically, we extract the attenuation features from the first, second, third, and fourth harmonics (i.e.,  $i = 1, 2, 3, 4$ ) given their prominent magnitude. Besides the attenuation ratio of harmonics at different levels, we also characterize the total distortion caused by vibration harmonics in all levels. In particular, we derive Total Harmonic Distortion (THD) and Harmonic-to-Total Ratio (HTR) from vital sign harmonics, which can be formulated as:

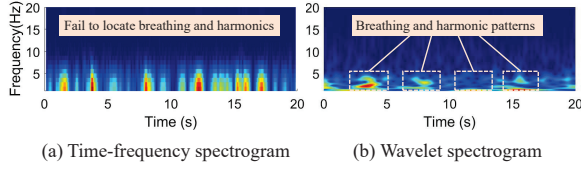
$$\text{THD}(x(t)) = \frac{\sqrt{\sum_i A_{h_i}^2(x(t))}}{A_b(x(t))}, \quad \text{HTR}(x(t)) = \frac{\sum_i P_{h_i}(x(t))}{P_t(x(t))}, \quad (5)$$

where  $P_t(\cdot)$  refers to total power of vital sign harmonics. Some selected features are shown in Figure 8, which indicates that our proposed attenuation features are effective and robust for differentiating users under vital sign variations across various XR scenarios. Beyond spatial features, we also find that the temporal patterns of vital sign vibration harmonics can provide additional features for user differentiation. Based on these observations, we include multiple breathing and heartbeat cycles in a data segment to reliably extract the proposed harmonic features. The extracted attenuation features will then be leveraged to generate high-fidelity representations for user authentication.

**Structural Feature Extraction.** Our system then extracts structural features, including cepstrogram and wavelet features, which characterize users' unique skull and internal facial structures.

(1) *Cepstrogram Feature Extraction.* Widely used in speech processing, cepstrogram features are able to characterize the properties of mechanical vibrations during propagation. As a kind of mechanical vibrations, vital sign harmonics propagate through the human skull and facial soft tissues, and carry unique structural properties.





**Figure 9: Breathing vibrations and harmonic patterns captured from time-frequency and wavelet spectrogram.**

Additionally, cepstrum is highly effective at detecting multi-path reflection and propagation effects, which are very common in the complex human skull and facial structures. The cepstrum feature extraction is formulated as:

$$C(q) = \text{STFT}^{-1} \left( \log |\text{STFT}(x(t))| \right), \quad (6)$$

where  $C(q)$  represents the cepstrum coefficients of the vibration signal  $x(t)$ ,  $q$  refers to the index of cepstrum features.

(2) *Wavelet Feature Extraction.* During the propagation of vital sign harmonics, the structural properties of human facial tissue could also affect the attenuation patterns of harmonic signals. In particular, they modify the frequency distributions by attenuating or amplifying different frequencies and induce time delays depending on the tissue's internal structural properties. Although Fourier Transform realizes frequency and harmonic analysis via a global spectrogram, its time-frequency resolution is fixed and cannot be dynamically adapted to different frequencies. In contrast, wavelet transform adaptively modifies the resolutions (e.g., high time resolutions on vibration patterns and high frequency resolutions on harmonics) to effectively depict the vibrations and their harmonics caused by complex internal facial structures. To extract wavelet features, we apply Continuous Wavelet Transform (CWT) to vital sign harmonics, which can be formulated as:

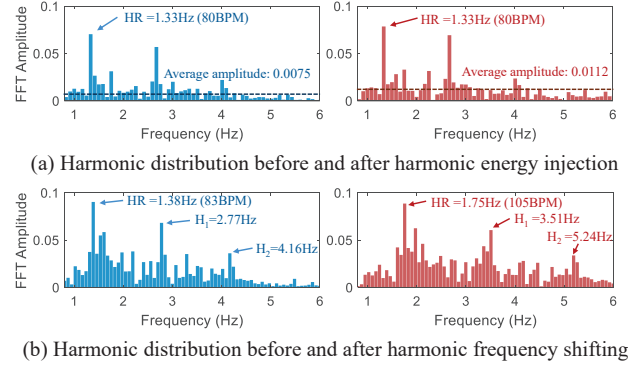
$$\text{CWT}_{a,\tau}(x(t)) = \langle x(t), \psi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int_L x(t) \psi\left(\frac{t-\tau}{a}\right) dt, \quad (7)$$

where  $\text{CWT}_{a,\tau}(\cdot)$  represents the function of CWT. where  $\text{CWT}_{a,\tau}(\cdot)$  represents the function of CWT.  $a$ ,  $\tau$ ,  $L$ , and  $\psi_{a,\tau}(t)$  refer to the frequency resolution, time resolution, signal length, and base function for transform, where we choose Morlet function given its generality on different frequency ranges. An example of the time-frequency and wavelet spectrogram on respiration vibrations and harmonics is shown in Figure 9, which indicates that the wavelet spectrogram achieves higher resolution in detecting respiration (e.g., 4 complete cycles in 20 seconds) and harmonic distributions.

### 5.3 Harmonic Augmentation

In practical usage scenarios, variations in vital sign signals may introduce subtle differences in vital sign harmonic patterns, and affect the biometric matching accuracy. Although robust harmonic features can be extracted to characterize unique skull structures and facial tissues, enriching the biometric patterns will further enhance the authentication performance of our proposed system. We accordingly develop a harmonic augmentation scheme with two techniques to improve the robustness of authentication.

**Harmonic Energy Injection.** Existing studies [60, 64] show that vital sign variations usually manifest as uncorrelated fluctuations similar to Gaussian white noise in respiration and heartbeat amplitudes. To simulate such variations on harmonic energy distribution, we process the Gaussian white noise via a bandpass filter



**Figure 10: Harmonic distributions before and after harmonic energy injection and harmonic frequency shifting.**

with the cut-off frequencies of respiration (0.1Hz ~ 0.5Hz) and heartbeat (0.8Hz ~ 3.0Hz). Then the processed noise is injected into the vital sign harmonics to simulate the energy distortion. By involving distorted vibrations for feature extraction and training, the authentication model will accommodate the amplitude variations of vital signs to enrich user biometric profiles. The harmonic energy injection process can be described as follows:

$$\hat{G}(t, \mu, \sigma^2) = A \cdot G(t, \mu, \sigma^2), \quad \bar{x}(t) = \hat{x}(t) + \hat{G}(t, \mu, \sigma^2) \quad (8)$$

0.1~0.5Hz/0.8~3.0Hz

where  $\hat{x}(t)$  represents the vital sign vibrations after mitigating motion artifacts via adaptive filtering.  $G(t, \mu, \sigma^2)$  denotes the Gaussian white noise with the average value  $\mu$ , standard deviation  $\sigma$ , and same length as  $\hat{x}(t)$ .  $\hat{G}(t)$  denotes the Gaussian noise after filtering with the cut-off frequency of breathing and heartbeat. We empirically set the average value  $\mu$  as 0 and the standard deviation  $\sigma$  as 0.05.  $A$  refers to the noise magnitude and is randomly selected from the range of [0, 0.005] for breathing and [0, 0.001] for heartbeat determined by empirical testing. For each vibration sample, we generate a series of noise samples for injection to ensure a comprehensive biometric enrichment. An example of harmonic distributions before and after energy injection is shown in Figure 10(a).

**Harmonic Frequency Shifting.** We further develop a harmonic frequency shifting scheme to simulate the distortion on vibration harmonic distributions caused by the duration variations of vital sign signals during practical XR usage. With the shifted patterns of vital sign harmonics being involved in feature extraction and training, the authentication model will enrich the user profiles under variations of vital sign periodicity. The harmonic frequency shifting can be formulated as:

$$\hat{x}(t) = \hat{x}(\gamma \cdot t), \quad \bar{x}(t) = \begin{cases} \hat{x}(t) \oplus \hat{x}(t') & 0 < \gamma < 1, \\ \hat{x}(t) & \gamma > 1, \end{cases} \quad (9)$$

where the shifting parameter  $\gamma$  is used to adjust the duration of vital sign vibration  $\hat{x}(t)$  and  $\oplus$  is the function of concatenation. In particular, the signal is compressed while  $0 < \gamma < 1$ . We refer to the re-scaled vibration as the concatenation of compressed vibration  $\hat{x}(t)$  and segments of the original vibration  $\hat{x}(t')$  defined within  $t \in [0, \gamma \cdot L]$  and  $t' \in [0, (1 - \gamma) \cdot L]$ , where  $L$  denotes the signal length. The signal is stretched while  $\gamma > 1$  with  $t \in [0, L]$ . An example of vibrations before and after harmonic frequency shifting is illustrated in Figure 10(b). With different  $\gamma$ , we generate a series of shifted vibration signals. These shifted signals will simulate

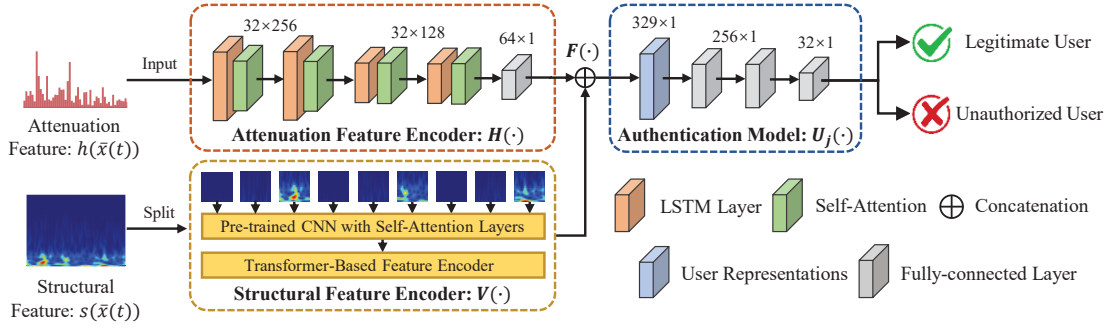


Figure 11: The overview of user authentication framework based on harmonic and spectral feature extraction.

the harmonic distortions caused by the variations of vital sign periodicity across various XR scenarios.

## 6 User Authentication Framework

**Attenuation Feature Encoding via Recurrent Network.** Given the unique attenuation properties of users' facial tissues, the vital sign harmonics exhibit distinct patterns. As a result, the attenuation features in vital sign harmonics manifest unique sequential patterns. Motivated by the effectiveness of Recurrent Neural Networks (RNN) in processing sequential inputs, we propose an attenuation feature encoder  $H(\cdot)$  based on Long Short-Term Memory (LSTM) and self-attention mechanism, as illustrated in Figure 11. Specifically, we involve two bidirectional LSTM layers with the output units of 256 and 128 to derive the sequential embeddings. To extract internal dependencies within harmonic patterns, we apply a self-attention layer for each LSTM layer with the query, key, and value equal to the output dimension of its connected LSTM layer. After passing the attenuation feature encoder, the embeddings will be concatenated with others to generate unique user representations.

**Transformer-based Structural Feature Encoding.** Architectures based on transformer, such as Vision Transformer (ViT) [30], have outperformed traditional deep learning models on image recognition tasks. Specifically, the multi-head self-attention mechanism in these models enable them to learn the correlations of different segments within the input sequences. Inspired by the architecture of ViT, we develop a transformer-based structural feature encoder  $V(\cdot)$  to extract user representations from the structural features. In particular, our transformer-based encoder splits the features into patches and encodes them with the temporal position index in the original patterns. We then embed these patches via the transformer-based encoder, including 4 convolutional layers with self-attention to extract user representations. The procedures of patch splitting and position encoding are shown in Figure 11.

**Optimization for Representation Extractor.** With attenuation and structural feature encoding, we combine the generated embeddings as the user representations. In particular, we optimize the parameters of the attenuation feature encoder  $H(\cdot)$  and structural feature encoder  $V(\cdot)$  to extract unique user embeddings. To recognize users via extracted representations, we build a user classifier  $P(\cdot)$  with two fully-connected layers. During training, we apply triplet loss and cross-entropy loss as loss functions to optimize model parameters. For triplet loss, it aims to maximize the difference among users with limited training data. It also ensures the harmonic augmentation does not make the system more vulnerable

to attacks by the adversaries, which can be formulated as:

$$F_p = \|F(x_b(t)) - F(x_p(t))\|_2, F_n = \|F(x_b(t)) - F(x_n(t))\|_2, \\ L_T = \frac{1}{N} \sum_{p,n} \max(F_p - F_n + \sigma, 0), \quad (10)$$

where  $x_b(t)$ ,  $x_p(t)$ , and  $x_n(t)$  denote the baseline, positive, and negative samples with user labels of  $y_b$ ,  $y_p$ , and  $y_n$  ( $y_b = y_p \neq y_n$ ).  $F(\cdot)$  refers to the combination of attenuation features  $h(x(t))$  and structural features  $s(x(t))$ .  $\sigma$  represents the margin between positive and negative pairs, which is empirically set as 0.1. To recognize the representations as correct users, we apply cross-entropy loss with triplet loss to optimize the encoder and user classifier, and the optimization procedure can be formulated as:

$$L_C = \frac{1}{N} \sum_{i=1}^N y_i \cdot \log(P(F(x_i(t))))), L = L_T + L_C, \quad (11)$$

where  $y_i$  and  $N$  denote the label of the  $i$ -th vibration input  $x_i(t)$  and the total number of vibration samples for training the feature encoder. Note that we recruit several participants and collect vibration samples from them as the training dataset. The user classifier  $P(\cdot)$  is only involved in training for extracting reliable user representations, and will not be utilized for user authentication.

**Optimization for User Authentication Model.** While training the user authentication model  $U(\cdot)$ , we fix the parameters of the harmonic feature encoders  $H(\cdot)$  and the spectral feature encoder  $V(\cdot)$ . For each legitimate user  $j$ , we collect a few samples for training a binary classifier  $U_j(\cdot)$  with two fully-connected layers to determine whether the input representations  $d$  belong to user  $j$  or not. The loss function used to optimize  $U_j(\cdot)$  can be described as:

$$L_{U_j} = \frac{1}{N} \sum_{i=1}^N y_i \cdot \log(U_j(d_i)), \quad (12)$$

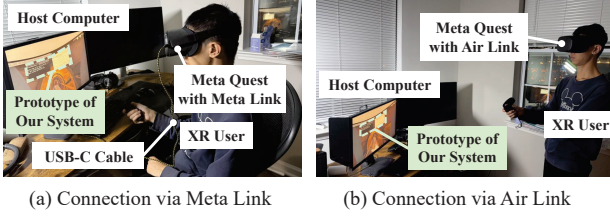
where  $y_i$  denotes the label of user representation  $d_i$ . Specifically, our system builds the classifier for each legitimate user using the training samples of the user and a few other people known by the system (e.g., users recruited by the developer), thus creating profiles for legitimate users and distinguishing them from adversaries.

## 7 Prototype Implementation

We build a prototype<sup>1</sup> of our system on Meta Quest to evaluate its on-device user enrollment and authentication. The headset is connected to a commercial desktop (i.e., with an Intel i9-13900K

<sup>1</sup>The demo of our designed system prototype can be found at our anonymous website: <https://sites.google.com/view/xrharmonics>





**Figure 12: Experimental setups of Meta Link with a USB-C cable and Meta’s Air Link for wireless connection.**

CPU and an NVIDIA GeForce RTX4090 GPU) via USB-C, which serves as the host computer, to run the XR and our system. The vibration processing, user enrollment, and authentication procedures are implemented on the host computer, emulating scenarios where the users’ biometrics are confined to their own XR devices. Besides the cable connection, we also evaluate the wireless connection between the headset and host computer (i.e., via Meta’s Air Link) for running our system. To avoid biometric leakage during data routing via the WiFi router, we leverage a direct WiFi connection by setting an ad-hoc network on the host computer. The data collection is realized using Meta SDK, with other ML functions implemented using Python 3.9.2 and TensorFlow 2.17. The experimental setups and examples of user authentication while watching an XR video are shown in Figure 12 and Figure 13. Specifically, the operation of our system includes three components:

(1) *Data Calibration & Feature Extraction.* This component is shared in both user enrollment and user authentication phases. Our system first mitigates the motion artifacts via adaptive filtering introduced in Section 5.1, which is realized using the functions in NumPy and SciPy. The calibrated data are then split into short segments (e.g., 3 seconds). During enrollment, our system augments the vibrations using our designed harmonic augmentation scheme and extracts representative features, which are implemented using the signal processing functions provided by SciPy.

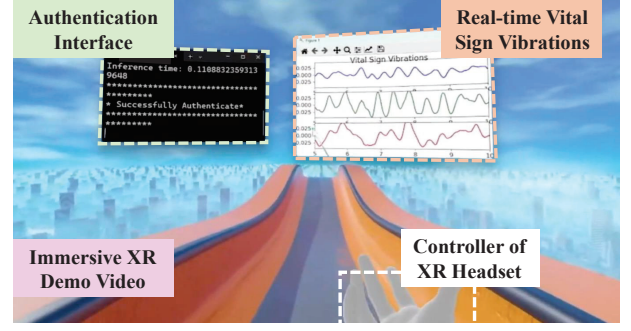
(2) *User Enrollment.* During user enrollment, our system automatically records the motion sensor readings over short periods (e.g., 2 minutes) from the users while they are using XR devices. Then our system extracts user biometrics using the representation extractor and builds the user profile via a deep-learning-based authentication model, both of which are implemented using the functions of TensorFlow/Keras. Note that the representation extractor is trained on a group of users known by our system and the enrollment phase does not need to re-train or fine-tune this model.

(3) *User Authentication.* With the profile of legitimate users, our system determines whether the current headset wearer is legitimate or not. For new vibration segments, our prototype calibrates the data, extracts features, and derives user representations. The system authenticates the wearer if the representation matches those in the profile; otherwise, it rejects the authentication attempts.

## 8 User Authentication Performance

### 8.1 Experimental Setups

**XR Headsets.** We evaluate the authentication performance of our system on Meta Quest and HTC Vive Pro Eye. Specifically, Meta Quest incorporates a motion sensor board (model 330-00193-03 1PASf8K) and HTC Vive Pro Eye utilizes an InvenSense MPU-6500



**Figure 13: A runtime example of our system for authentication under XR scenarios (e.g., watching an XR demo video).**

sensor board with an accelerometer and a gyroscope. During the data collection, we leverage the function `ovr_GetTrackingState()` from Meta SDK to access the motion sensor of Meta Quest. For HTC Vive Pro Eye, we use the built-in lighthouse application to extract the motion sensor readings from the headset. For both headsets, we set the sampling rates of the motion sensors to 1000Hz.

**XR Scenarios Involved in Data Collection.** We collect vital sign vibrations under four common XR scenarios. (1) *Sitting and Watching a Demo Video:* Participants wearing the XR headset sit and watch an immersive XR video for one minute, and remain stationary. (2) *Standing and Watching a Demo Video:* Participants stand and watch a demo video for one minute. Compared to the first scenario, the participants exhibit subtle involuntary movements of their heads or bodies. (3) *Using Controllers to Browse Application Store:* Participants use controllers for browsing motions, such as arm raising and lowering, which slightly affect the headset motion sensors. (4) *Head rotation in the Virtual Environment:* Participants rotate their head from left to right, which causes significant motion sensor fluctuations due to instant changes in acceleration. To ensure accurate authentication, we applied our adaptive filtering scheme to minimize the motion artifacts on vibration signals. Moreover, the four XR scenarios are performed under four different real-world environments (e.g., four offices with different layouts and furniture). By involving different XR scenarios and environments, we aim to validate that our system will effectively authenticate users under practical contexts and various physical locations.

**Participants and Data Collection.** We collect the vital sign vibrations from 52 participants aged from 18 to 45. Specifically, we involve 45 participants (30 males and 15 females) to collect the data from Meta Quest, while the dataset of HTC Vive Pro includes 27 participants (17 males and 10 females). Before the data collection, the participants are provided with a consent document, including a detailed description of the procedure, risks, and potential discomforts. They have the option to accept or decline to collect the data. In addition, our research team signs a confidential agreement with each participant for biometric protection, with data of each participant labeled as a pseudo-identity during model training and testing. The data collection procedure has been approved by our university’s Institutional Review Boards (IRB). Each participant wears the headset for data collection under all four scenarios. Note that prior studies [36, 39, 72] for authentication are usually evaluated with less than 30 users, thus the number of users for evaluation is in line with existing works. The collected vibrations are then split into

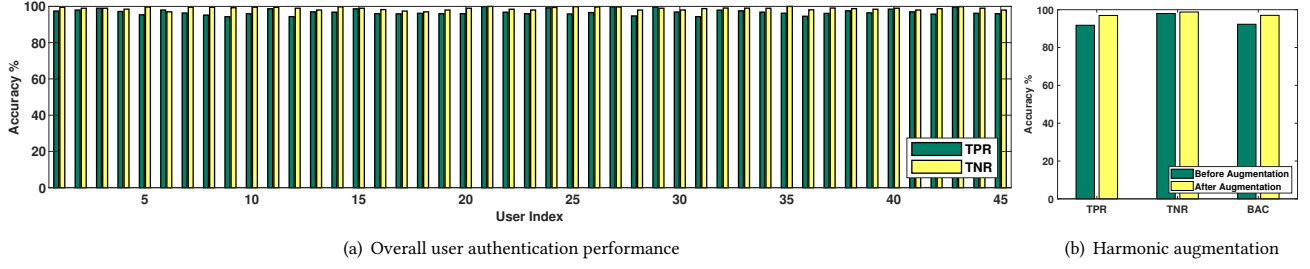


Figure 14: Overall user authentication performance and evaluation on our proposed harmonic augmentation scheme.

Table 2: User authentication performance of our system with (w/) and without (w/o) Adaptive Filtering (AF).

|            | Standing |        | Moving arms |        | Head Rotation |        |
|------------|----------|--------|-------------|--------|---------------|--------|
|            | w/o AF   | w/ AF  | w/o AF      | w/ AF  | w/o AF        | w/ AF  |
| <b>TPR</b> | 66.29%   | 96.14% | 41.67%      | 95.79% | 24.29%        | 94.78% |
| <b>TNR</b> | 90.21%   | 98.45% | 84.49%      | 97.27% | 77.45%        | 97.94% |
| <b>BAC</b> | 70.24%   | 96.56% | 42.96%      | 96.11% | 26.65%        | 95.55% |

3-second segments. In total, we collect 5,467 and 3,389 vibration segments from Meta Quest and HTC Vive Pro Eye.

**Evaluation Metrics.** We employ the following metrics to evaluate our system. (1) *True Positive Rate (TPR)*: The percentage of samples from legitimate users that are correctly authenticated. (2) *True Negative Rate (TNR)*: The percentage of samples from unauthorized users that are accurately rejected. (3) *Balanced Accuracy (BAC)*: a weighted average of TPR and TNR, reflecting the authentication accuracy under imbalanced positive and negative samples. Higher TPRs and TNRs (i.e., low FPRs and FNRs) indicate higher accuracy in authenticating legitimate users and rejecting adversaries, and BAC quantifies the overall performance of our system.

## 8.2 Overall User Authentication Performance

We utilize and combine the harmonics from respiration and heartbeat collected from Meta Quest under four scenarios to evaluate the overall authentication performance of our system. Specifically, we take turns setting each of 45 participants as the legitimate user and the remaining as adversaries. The vibration samples are separated with a ratio of 8:2 for training and testing the authentication model. The TPR and TNR of each user are shown in Figure 14(a). The results show that our system achieves TPRs of more than 94.14% and TNRs of more than 97.79% for all 45 participants. High TPRs and TNRs demonstrate that our system accurately authenticate users via vital sign harmonics. We also summarize the average TPR, TNR, and BAC before and after harmonic augmentation, which are shown in Figure 14(b). Without augmentation, the average TPR, TNR, and BAC of our system are 91.77%, 97.98%, and 92.29%. After applying harmonic augmentation, the average TPR, TNR, and BAC achieve more than 96.95%, 98.77%, and 97.01%. The improvements validate that our proposed harmonic augmentation scheme will facilitate accurate authentication under vital sign variations.

## 8.3 Study of Impact Factors

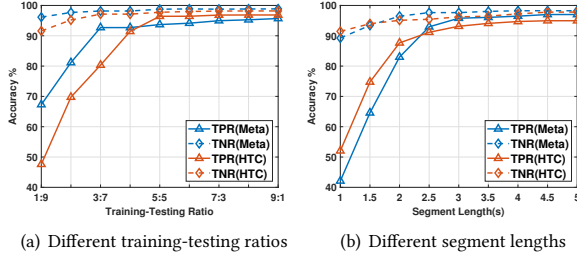
**Impact of Motion Artifact.** To evaluate the motion artifact mitigation scheme via adaptive filtering, we conduct experiments on Meta Quest with 20 participants involved. Specifically, we collect

and combine the breathing and heartbeat harmonics under the scenarios of standing, using controllers, and head rotation. To evaluate our system, we split the collected samples into training and testing sets with a ratio of 8:2. The average TPR, TNR, and BAC with and without motion artifact mitigation are summarized in Table 2. Under the head rotation scenario, the performance is significantly degraded by motion artifacts, with average TPRs, TNRs, and BACs below 24.29%, 77.45%, and 26.65%. In the standing scenario, the performance is also affected by body motions, with TPR, TNR, and BAC less than 66.29%, 90.21%, and 70.24%. After mitigating motion artifacts, the TPR, TNR, and BAC achieve 94.78%, 97.94%, 95.55% under head rotation and 96.14%, 98.45%, 96.56% in the standing scenario. Prominent improvements in authentication accuracy validate that our method can effectively mitigate motion artifacts, thus enhancing system deployment under general XR scenarios.

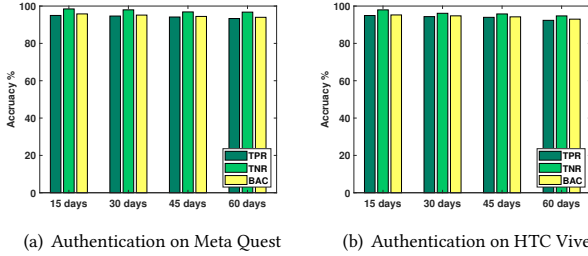
**Impact of Sample Length for Enrollment.** Although increasing the number of vibration samples for user enrollment could improve the authentication accuracy of our system, it introduces additional efforts on data collection. To evaluate the enrollment cost of our system, we collect the vibration samples on Meta Quest and HTC Vive Pro Eye from 20 users under the sitting scenario. The TPR and TNR with different training-testing ratios are summarized in Figure 15(a). For Meta Quest, our system achieves TPR and TNR of more than 92.70% and 98.17% with the training-testing ratio of 4:6. For HTC Vive Pro Eye, the TPR and TNR exceed 91.44% and 97.09% under the same training-testing ratio. High authentication accuracy across training-testing ratios validates the system effectiveness with low enrollment costs. The TPR and TNR with different segment lengths are shown in Figure 15(b). For Meta Quest, the TPR and TNR achieve 95.79% and 97.64% with the frame length of 3 seconds. For HTC Vive Pro Eye, the TPR and TNR also reach 94.65% and 97.33% with the same frame length. High authentication accuracy with short-time data collection validates the low cost of our system for user enrollment.

## 8.4 Long-Term Authentication Study

Morphological variations in users' vital signs and harmonic signals over a long period could downgrade the robustness of our system. To evaluate the long-term performance of our system, we conduct experiments by collecting and combining breathing and heartbeat harmonics using Meta Quest and HTC Vive Pro Eye from 15 users under all four XR scenarios. Specifically, the data is collected 5 times within 60 days, with each collection across 15 days. We use the vibrations collected on day 0 to build the authentication model and data from other days (e.g., 15, 30, 45, 60) to evaluate the performance.



**Figure 15: Authentication performance of our system with different training-testing ratios and lengths of samples.**

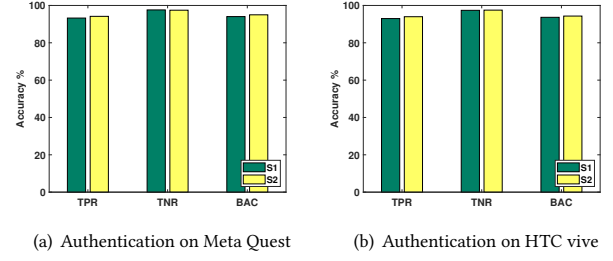


**Figure 16: Authentication performance on Meta Quest and HTC Vive Pro Eye across different time intervals.**

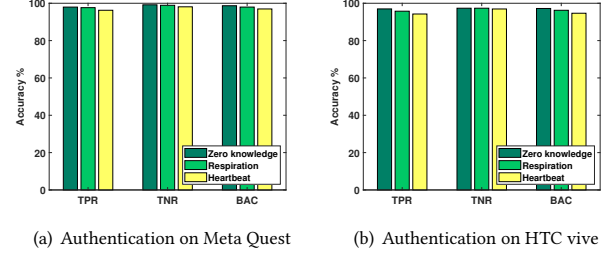
We take turns selecting each user as the legitimate user and the average TPRs, TNRs, and BACs are shown in Figure 16. For Meta Quest, our system attains TPR, TNR, and BAC of 94.97%, 98.45%, and 95.79% while the authentication performance is evaluated across 15 days. After 60 days, the TPR, TNR, and BAC still remain 93.32%, 96.75%, 93.97%, which demonstrates that our system achieves long-term authentication on Meta Quest. For HTC Vive Pro Eye, our system achieves TPR, TNR, and BAC of over 94.93%, 97.95%, and 95.24% for authentication across 15 days. Across 60 days, our system still remains high TPR, TNR, and BAC of over 92.35%, 94.72% and 92.98%. High authentication accuracy across long time intervals validates the system effectiveness on user authentication under long-term usage. Additionally, our system can leverage adaptive training by periodically updating user profiles to remain effective as the vital sign harmonics can be passively collected from users.

### 8.5 Cross-Session Authentication Performance

To evaluate the system performance across different XR scenarios, we conduct experiments by collecting and combining vital sign harmonics using Meta Quest by and HTC Vive Pro Eye from 10 users under two separate sessions: (1) *Session 1 (S1)*: The users sit still and scan an XR webpage. (2) *Session 2 (S2)*: The users stand still and watch an XR spacewalk video. We then take turns selecting each user as the legitimate user to evaluate the performance and the average TPRs, TNRs, and BACs are shown in Figure 17. For Meta Quest, the TPR, TNR, and BAC achieve 94.17%, 97.46%, and 94.99% with S1 for enrollment and S2 for authentication. Using S2 for enrollment, the TPR, TNR, and BAC with S1 for authentication also reach 93.25%, 97.61%, and 94.04%. For HTC Vive, the TPR, TNR, and BAC achieve 93.97%, 97.50%, and 94.32% with S1 for enrollment and S2 for authentication. Using S2 for enrollment, the TPR, TNR, and BAC also achieve 92.98%, 97.39%, and 93.65%. High authentication accuracy across different XR sessions validates that our system



**Figure 17: Authentication performance on Meta Quest and HTC Vive Pro Eye across different XR sessions.**



**Figure 18: Authentication performance on Meta Quest and HTC Vive Pro Eye against potential attacks.**

realizes robust authentication in different XR scenarios via biometric augmentation. Additionally, we evaluate the authentication performance by reproducing the feature design in FaceReader [69] and the TPR, TNR, and BAC are 69.05%, 82.12%, and 72.73%. These results further validate that our proposed biometric features remain more consistent under different physiological states of users.

### 8.6 Evaluation of Authentication Latency

Real-time authentication is essential for enhancing the user experience in practical systems. To validate that our system achieves real-time authentication, we evaluate the average time cost of different components on Meta Quest with the setup described in Section 7. Specifically, we measure the average time latency of vibration calibration, feature extraction, user enrollment, and authentication, of 10 participants while they are rotating their heads. Particularly, the vibration calibration and feature extraction have average time costs of 52.44ms and 445.29ms for 3-second samples. For user enrollment, the average time latency is 14.35s, including vibration segmentation, biometric extraction, and model training. For authentication, the average time cost is 12.45ms. Combined with the vibration length (e.g., 3 seconds), our system takes 3.51s to authenticate users, which is much shorter compared to existing approaches (e.g., ~7.0s in SoundLock [72], ~11.2s in BlinKey [71], and ~6.0s in GaitLock [61]) and indicates that our system can realize real-time authentication. We also measure the data routing latency while users leverage Meta's Air Link to access the host computer for running our system. The average latency is less than 3.0ms, which validates the real-time authentication with the wireless connection.

## 9 Robustness To Spoofing Attack

**Robustness to Attack Scenario I.** In this scenario, the adversary attempts to bypass our system using the vital sign harmonics from their own or the recruited users. In particular, we collect the vital



**Table 3: The frequency response of different types of commercial vibration motors.**

| Category                       | Vibration Motor            | Frequency Range |
|--------------------------------|----------------------------|-----------------|
| Eccentric Rotating Mass (ERM)  | Precision Brushed ERM [15] | 30Hz ~ 500Hz    |
|                                | NFP-P0615-4302 [4]         | 140Hz ~ 190Hz   |
| Linear Resonant Actuator (LRA) | Precision LRA [14]         | 150Hz ~ 205Hz   |
|                                | Vybronic VL32158H-L25 [16] | 90Hz ~ 110Hz    |
| Electromagnetic Motors (EM)    | TOAUTO TO302S1 [12]        | 50Hz ~ 60Hz     |
|                                | ATO-AVB-TB1002S2 [5]       | 50Hz ~ 60Hz     |

sign harmonics from 10 participants on Meta Quest and HTC Vive Pro Eye with each selected as the legitimate user. We then collect the data from the other 10 users and combine them with those from legitimate users for testing. The evaluation results are shown in Figure 18. For Meta Quest, our system achieves TPR, TNR, and BAC of more than 97.95%, 99.17% and 98.67%. For HTC Vive, the system remains high authentication accuracy with TPR, TNR, and BAC of 96.97%, 97.35%, and 97.21%. High authentication accuracy in recognizing legitimate users validates the robustness of our proposed feature design. Additionally, the system effectively rejects potential adversaries, indicating that the system can accurately differentiate legitimate users from unauthorized users with our approach.

**Robustness to Attack Scenario II.** In this scenario, the adversaries have prior knowledge of the respiration or heartbeat (e.g., patterns, rates) of legitimate users. Specifically, we select 5 participants as legitimate users and collect their vital sign harmonics from Meta Quest and HTC Vive Pro Eye to train the authentication model. For each legitimate user, we select 2 users with similar demographic (e.g., gender, age) and body measurements (e.g., weight, height, body fat ratio) as adversaries. We then use a metronome to ensure that the adversaries and legitimate users have similar respiration rates and the adversaries selected share similar resting heart rates with legitimate users. During testing, we collect the vital sign harmonics from the adversaries and combine them with those from legitimate users. The average TPRs, TNRs, and BACs on Meta Quest are shown in Figure 18(a). Specifically, our system achieves TPR, TNR, and BAC of 97.65%, 98.92%, and 97.96%, which validates the system's robust authentication against spoofing attacks. The performance on HTC Vive Pro Eye is shown in Figure 18(b), with TPR, TNR, and BAC of over 95.75%, 97.35%, and 96.24%. For the scenario that the adversary has prior knowledge of users' heartbeat, our system achieves TPR, TNR, and BAC of 96.25%, 98.10%, and 96.95%, which validates the effectiveness on rejecting adversaries. For HTC Vive, the TPR, TNR, and BAC also reach 94.25%, 96.94%, and 94.66%. Consistently high accuracy under attack validates that our system effectively recognizes legitimate users and rejects adversaries by capturing users' unique skull and facial tissue properties.

**Analysis of Vibration Replay Attack.** The adversary may consider using programmable vibration motors to replay vibrations of a user's vital signs to spoof our system. However, reproducing vibrations requires a vibration motor that can generate extremely low-frequency vibrations (e.g.,  $\leq 3.0$ Hz). The frequency responses of common vibration motors are summarized in Table 3. We find that none of the commercial vibration motors can produce vibrations within the frequency ranges of vital signs. Specifically, the vibration motors based on eccentric rotating mass normally generate vibrations higher than 30Hz and the other vibration motors

**Table 4: Efficient to Use (EU), Physically Effortless (PE), Memorywise Effortless (ME), Nothing to Carry (NC), Resilient to Spoofing (RS), and Resilient to Observation (RO) on our system and existing XR authentication systems.**

| Category                | Method               | Name                       | EU                  | PE | ME | NC | RS | RO |
|-------------------------|----------------------|----------------------------|---------------------|----|----|----|----|----|
| Knowledge-based         | PIN                  | PIN [32, 68]               | O                   | ×  | ×  | ✓  | ×  | ×  |
|                         |                      | Voice-based PIN [66]       | ×                   | O  | ×  | ✓  | ×  | ×  |
|                         |                      | gTalker [42, 43]           | O                   | O  | ×  | ✓  | ×  | ✓  |
|                         | Password             | Swipe Pattern [32, 56]     | O                   | ×  | ×  | ✓  | ×  | ×  |
| Token-based             | QR Code              | Glass OTP [27]             | N                   | O  | ✓  | ×  | ✓  | ✓  |
| Physiological Biometric | Eye                  | Iris & Periocular [26, 34] | N                   | ✓  | ✓  | ✓  | ×  | ✓  |
|                         |                      | Visual Stimulus [70]       | ✓                   | ✓  | ✓  | ✓  | ×  | ✓  |
|                         |                      | Head                       | Brain Password [45] | ×  | ✓  | ✓  | ×  | O  |
|                         |                      | SkullConduct [59]          | ✓                   | ✓  | ✓  | ✓  | ×  | ✓  |
|                         | Muscle               | ElectricAuth [29]          | O                   | ✓  | ✓  | ×  | ✓  | ✓  |
|                         | Behavioral Biometric | Eye                        | Oculock [47]        | ×  | ×  | ✓  | ×  | O  |
| Gait/Head               |                      |                            | Virtual Scene [54]  | O  | ✓  | ✓  | ✓  | ×  |
|                         |                      | Headbanger [40]            | ×                   | ×  | ×  | ✓  | O  | ×  |
|                         |                      | GaitLock [61]              | O                   | ×  | ✓  | ✓  | ✓  | O  |
| Hand                    |                      | Glass Interactions [28]    | O                   | O  | ✓  | ✓  | ✓  | ×  |
|                         |                      | Throw Trajectory [35]      | O                   | ×  | ✓  | ✓  | O  | ×  |
| Multi-factor            | Brain                | EEG Signal [41]            | ×                   | ×  | ×  | ×  | ✓  | ✓  |
|                         | Eye                  | BlinkKey [71]              | ×                   | O  | ×  | ✓  | ✓  | ✓  |
|                         | Gait/Head            | GlassGesture [67]          | N                   | ×  | ×  | ✓  | ✓  | ✓  |
|                         | Hand                 | RubikBiom [50]             | O                   | ×  | ×  | ✓  | ✓  | ✓  |
| Physical & Behavioral   | Harmonics            | Ours                       | ✓                   | ✓  | ✓  | ✓  | ✓  | ✓  |

✓: fulfills criterion    ×: does not fulfill criterion    O: quasi-fulfills criterion    N: not provided

based on linear resonant actuators and electromagnetic motors are designed to generate vibrations over 50Hz. It is infeasible for an adversary to replicate vital sign harmonics using these vibration motors, even if they have access to users' breathing and heartbeat.

## 10 Related Work

Current XR systems (e.g., Meta Quest) mainly adopt knowledge-based methods, such as passwords and PIN numbers, to perform user authentication. However, these methods do not align well with the unique XR input interfaces relying on gestures and voice input. In addition, the gestures have been proven vulnerable to observation/shoulder-surfing attacks [21, 48, 51]. A few studies attempt to improve the security and usability of password/PIN-based authentication [32, 66, 68]. Specifically, spoken PINs with obfuscated digits [42, 43, 66] are designed to prevent password leakage via voice eavesdropping. Besides, randomly generated tokens [13] and QR codes [10] on a second device (e.g., smartphone or wearable) are employed. However, these methods require additional user effort and could interrupt the immersive XR experience.

Recently, behavioral and physiological biometrics have been investigated to authenticate users. For example, unique behaviors embedded in hand gestures [28, 35], head rotations [40, 54], and eye movements [47]. However, these biometrics rely on active user interactions, limiting their application scenarios. Researchers have also exploited physiological biometrics for authentication, such as iris patterns from cameras [26, 34], muscle properties from stimulation sensors [29], and brainwave patterns from brainwave sensors [45]. Although these methods are effortless, they require extra hardware or specialized sensors that are unavailable on mainstream headsets. Moreover, multi-factor schemes integrate multiple authentication factors [50, 67, 71], such as passwords/PINs and behavioral biometrics, to improve security. They ensure a higher degree of security against observation attacks, but inherit the usability problem in each factor (e.g., active gestures). The weaknesses of existing XR authentication schemes are summarized in Table 4.

A recent study (i.e., FaceReader [69]) designs a privacy attack that reconstructs breathing and heartbeat patterns from facial vibrations.

Although it shows a case study of user tracking, fundamental differences exist in terms of objectives, scope, and technical challenges. FaceReader is a privacy attack, while ours is a benign and end-to-end authentication system. FaceReader trains with a fixed dataset and cannot detect users who are not involved in training. Under the context of authentication, FaceReader cannot recognize unknown adversaries/spoofers. In contrast, our system detects spoofers with a complete authentication pipeline. Secondly, our system introduces novel harmonic features to characterize unique human skull/facial tissue properties with greater long-term consistency, whereas FaceReader leverages raw vital sign pattern to realize short-term user tracking. Thirdly, our system realizes robust authentication under vital sign variations across XR scenarios via harmonic augmentation based on energy injection and frequency shifting. In contrast, FaceReader re-identifies users under similar conditions.

## 11 Discussion of Biometric Protection

**Analysis of False Positive Cases.** By scrutinizing the misclassified cases during authentication, we find that most false positive cases arise in two particular scenarios: (1) at the beginning or end of usage sessions when device placement or removal causes instability, and (2) when wearing position slightly shifts across different sessions. Both scenarios will introduce subtle alignment variations between the XR headsets and the users' faces, and then cause distribution shifts from the extracted features to the enrolled user profiles. To address this concern, we plan to explore adaptive training strategies in future work, which improve authentication robustness against cross-session variations during practical deployment.

**Protecting Biometrics from Other Applications.** While running our system, motion sensor readings with vital sign harmonics can also be accessed by other applications, posing risks of biometric leakage if a malicious application is running in the back-end. To mitigate this risk, a possible solution is to filter out vital sign harmonics (e.g., band-pass filtering) for low-priority applications that do not require vital sign data (e.g., gaming, social media) as they maintain full functionality using filtered data. High-priority access can be granted to our system, allowing it to receive motion sensor data for authentication. This solution not only ensures the functioning of other applications but also strengthens protection against attacks that eavesdrop on motion sensors [69].

**On-device Authentication.** The biometrics of vital sign harmonics may also be leaked during data transmission, such as sending the data to a cloud or edge server for building the authentication model. To mitigate these risks, we have explored on-device execution for our system, with our prototype introduced in Section 7. This allows users to store their biometrics and run the models on their XR headsets locally without data transmission. Multi-party computation can also be adopted, where multiple servers jointly generate the complete user representations to prevent user biometric leakage. Our system can also employ biometric encryption techniques (e.g., homomorphic encryption) to prevent biometric leakage on personal devices or cloud servers with insecure or corrupted memory.

## 12 Conclusion

In this paper, we propose the first effortless and inbuilt user authentication system for XR headsets by harnessing the unique skull and

facial tissue biometrics from vital sign harmonics. Different from prior works, our system passively collects vital sign harmonics during regular XR usage, without requiring any active user inputs and additional hardware. We develop an adaptive filtering method that dynamically mitigates the body motion artifacts in vital sign harmonics. Based on the ratios among different harmonic components, we propose harmonic features that are non-volatile under vital sign variations to depict users' unique skull structures and facial tissue properties. In addition, we design a harmonic augmentation strategy based on energy injection and frequency shifting to simulate the physiological variations across XR contexts. We further employ transformer models and self-attention mechanisms to create high-fidelity user representations. Extensive evaluations validate the effectiveness of our system in authenticating legitimate users and rejecting adversaries in diverse XR scenarios.

## Acknowledgment

This work was partially supported by the National Science Foundation Grants CNS2201465, CNS2329278, CNS2120396, CNS2120276, CNS2145389, CNS2329279, OAC2139358, CCF2211163, DGE2414365, DGE2414366, and DGE2414367.

## References

- [1] 2023. 24+ Augmented Reality Stats (2024-2028). <https://explodingtopics.com/blog/augmented-reality-stats>.
- [2] 2023. The State of XR in Manufacturing and Industrial 2023. <https://www.xrtoday.com/mixed-reality/the-state-of-xr-in-manufacturing-and-industrial-2023/>.
- [3] 2023. What is mixed reality? <https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality>.
- [4] 2024. 6mm Small Vibration Device – 15mm Type Model NFP-P0615-4302. <https://nfpshop.com/product/6mm-vibration-motor-15mm-type-small-vibration-motor-model-nfp-p0615-4302>.
- [5] 2024. ATO-AVB-TB1002S2 100W AC Vibration Motor, 1 Phase, 2 Pole. <https://www.ato.com/100w-asynchronous-vibration-motor>.
- [6] 2024. Augmented and Virtual Reality in Social Media Marketing. <https://wearstudio.com/augmented-and-virtual-reality-in-social-media-marketing/>.
- [7] 2024. Extended Reality Research and Technology. <https://www.qualcomm.com/research/extended-reality>.
- [8] 2024. How VR and AR are Changing the Face of Entertainment and Beyond. <https://www.tdk.com/en/tech-mag/past-present-future-tech/ar-vr-mr>.
- [9] 2024. Immersive Learning: How Virtual And Augmented Reality Are Transforming Higher Education. <https://elearningindustry.com/immersive-learning-how-virtual-and-augmented-reality-transforming-higher-education/>.
- [10] 2024. Install the Meta Quest mobile app on your phone. <https://www.meta.com/help/quest/articles/getting-started/getting-started-with-quest-2/install-meta-quest-mobile-app/>.
- [11] 2024. Manage two-factor authentication for your Apple ID from Apple Vision Pro. <https://support.apple.com/zh-cn/guide/apple-vision-pro/tan9757ade3/visionos>.
- [12] 2024. TOAUTO Concrete Vibrator Vibration Motor 30W Single Phase AC 110V 3600rpm Aluminum Alloy Case Vibrating Motor for Shaker Table. <https://www.sears.com/toauto-concrete-vibrator-vibration-motor-60w-single-phase-ac-110v-3600rpm-aluminum-alloy-case-vibrating-motor-for-p-A122459019>.
- [13] 2024. Two-Factor Authentication on VRChat. <https://docs.vrchat.com/docs/setup-2fa>.
- [14] 2024. Vibration Motor Comparison Guide. <https://www.precisionmicrodrives.com/ab-028>.
- [15] 2024. Vibration Motors – ERMs and LRAs. <https://www.precisionmicrodrives.com/vibration-motors-erms-and-lras>.
- [16] 2024. VL32158H-L25. <https://www.vybronic.com/linear-lra-vibration-motors/vl32158h-l25>.
- [17] 2024. What is augmented reality or AR? <https://dynamics.microsoft.com/en-us/mixed-reality/guides/what-is-augmented-reality-ar/>.
- [18] 2024. What is Virtual Reality? <https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html>.
- [19] 2024. XRHealth Virtual Clinic: At-Home Virtual Reality Therapy. <https://www.xr.health/>.
- [20] Hamed Abdi, David Sánchez-Molina, Silvia García-Vilana, and Vafa Rahimi-Movaghar. 2024. Revealing the role of material properties in impact-related

- injuries: Investigating the influence of brain and skull density variations on head injury severity. *Heliyon* 10, 8 (2024).
- [21] Yasmeen Abdurabou, Sheikh Radiah Rivu, Tarek Ammar, Jonathan Liebers, Alia Saad, Carina Liebers, Uwe Gruenefeld, Pascal Knierim, Mohamed Khamis, Ville Makela, Stefan Schneegass, and Florian Alt. 2022. Understanding Shoulder Surfer Behavior and Attack Patterns Using Virtual Reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) (AVI 2022). Association for Computing Machinery, New York, NY, USA, Article 15, 9 pages. <https://doi.org/10.1145/3531073.3531106>
  - [22] Pierre G Agache, C Monneur, Jean Luc Leveque, and Jean De Riga. 1980. Mechanical properties and Young's modulus of human skin in vivo. *Archives of dermatological research* 269 (1980), 221–232.
  - [23] Hind Alrubaish and Nazar Saqib. 2022. Your Vital Signs as Your Password? In *Recent Advances in Biometrics*, Muhammad Sarfraz (Ed.). IntechOpen, Rijeka, Chapter 10. <https://doi.org/10.5772/intechopen.104783>
  - [24] Abdullah Al Arafat, Zhishan Guo, and Amro Awad. 2021. VR-Spy: A Side-Channel Attack on Virtual Key-Logging in VR Headsets. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 564–572. <https://doi.org/10.1109/VR50410.2021.00081>
  - [25] Bekir Bediz, H Nevzat Özgüven, and Feza Korkusuz. 2010. Vibration measurements predict the mechanical properties of human tibia. *Clinical biomechanics* 25, 4 (2010), 365–371.
  - [26] Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper. 2020. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. *Image and Vision Computing* 104 (2020), 104007.
  - [27] Pan Chan, Tzipora Halevi, and Nasir Memon. 2015. Glass otp: Secure and convenient user authentication on google glass. In *Financial Cryptography and Data Security: FC 2015 International Workshops, BITCOIN, WAHC, and Wearable, San Juan, Puerto Rico, January 30, 2015, Revised Selected Papers*. Springer, 298–308.
  - [28] Jagmohan Chauhan, Hassan Jameel Asghar, Mohamed Ali Kaafar, and Anirban Mahanti. 2014. Gesture-based continuous authentication for wearable devices: the google glass case. *arXiv preprint arXiv:1412.2855* (2014).
  - [29] Yuxin Chen, Zhuolin Yang, Ruben Abbou, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2021. User authentication via electrical muscle stimulation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
  - [30] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv:2010.11929* [cs.CV]
  - [31] Mostafa Fatemi, Armando Manduca, and James F Greenleaf. 2003. Imaging elastic properties of biological tissues by low-frequency harmonic vibration. *Proc. IEEE* 91, 10 (2003), 1503–1519.
  - [32] Ceenu George, Mohamed Khamis, Emanuel von Zeszschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. NDSS.
  - [33] 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](https://doi.org/10.1007/978-3-642-04898-2_327)
  - [34] Sehee Kim and EuiChul Lee. 2018. Periocular biometric authentication methods in head mounted display device. *Int. J. Eng. Technol* 7, 3 (2018).
  - [35] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-driven biometric authentication of users in virtual reality (VR) environments. In *MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I* 25. Springer, 55–67.
  - [36] Sunwoo Lee, Wonsuk Choi, and Dong Hoon Lee. 2021. Usable User Authentication on a Smartwatch using Vibration. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (Virtual Event, Republic of Korea) (CCS '21). Association for Computing Machinery, New York, NY, USA, 304–319. <https://doi.org/10.1145/3460120.3484553>
  - [37] Roman Lewandowski and B Chorańczyczewski. 2010. Identification of the parameters of the Kelvin–Voigt and the Maxwell fractional models, used to modeling of viscoelastic dampers. *Computers & structures* 88, 1–2 (2010), 1–17.
  - [38] Feng Li, Jiayi Zhao, Huan Yang, Dongxiao Yu, Yuanfeng Zhou, and Yiran Shen. 2023. VibHead: An Authentication Scheme for Smart Headsets through Vibration. *arXiv:2306.17002* [cs.CR]
  - [39] Jingjie Li, Kassem Fawaz, and Younghyun Kim. 2019. Velody: Nonlinear Vibration Challenge-Response for Resilient User Authentication. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) (CCS '19). Association for Computing Machinery, New York, NY, USA, 1201–1213. <https://doi.org/10.1145/3319535.3354242>
  - [40] Sugang Li, Ashwin Ashok, Yanyong Zhang, Chenren Xu, Janne Lindqvist, and Macro Gruteser. 2016. Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–9.
  - [41] Sukun Li, Sonal Savaliya, Leonard Marino, Avery M. Leider, and Charles C. Tappert. 2019. Brain Signal Authentication for Human-Computer Interaction in Virtual Reality. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. 115–120. <https://doi.org/10.1109/CSE/EUC.2019.00031>
  - [42] Yan Li, Yao Cheng, Yingjiu Li, and Robert H Deng. 2017. What you see is not what you get: Leakage-resilient password entry schemes for smart glasses. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*. 327–333.
  - [43] Yan Li, Yao Cheng, Weizhi Meng, Yingjiu Li, and Robert H Deng. 2020. Designing leakage-resilient password entry on head-mounted smart wearable glass devices. *IEEE Transactions on Information Forensics and security* 16 (2020), 307–321.
  - [44] Jonathan Liebers and Stefan Schneegass. 2020. Gaze-based Authentication in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) (ETRA '20 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 41, 2 pages. <https://doi.org/10.1145/3379157.3391421>
  - [45] Feng Lin, Kun Woo Cho, Chen Song, Wenya Xu, and Zhanpeng Jin. 2018. Brain password: A secure and truly cancelable brain biometrics for smart headwear. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 296–309.
  - [46] Shiqing Luo, Anh Nguyen, Hafsa Farooq, Kun Sun, and Zhisheng Yan. 2024. Eavesdropping on Controller Acoustic Emanation for Keystroke Inference Attack in Virtual Reality. <https://doi.org/10.14722/ndss.2024.24100>
  - [47] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenya Xu, and Zhisheng Yan. 2020. OcuLock: Exploring human visual system for authentication in virtual reality head-mounted display. In *2020 Network and Distributed System Security Symposium (NDSS)*.
  - [48] Anindya Maiti, Murtuza Jadliwala, and Chase Weber. 2017. Preventing shoulder surfing using randomized augmented reality keyboards. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 630–635. <https://doi.org/10.1109/PERCOMW.2017.7917636>
  - [49] Marzena Malińska, Krystyna Zużewicz, Joanna Bugajska, and Andrzej Grabowski. 2015. Heart rate variability (HRV) during virtual reality immersion. *International Journal of Occupational Safety and Ergonomics* 21 (2015), 47–54. <https://api.semanticscholar.org/CorpusID:18043598>
  - [50] Florian Mathis, Hassan Ismail Fawaz, and Mohamed Khamis. 2020. Knowledge-driven biometric authentication in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–10.
  - [51] Florian Mathis, Joseph O'Hagan, Mohamed Khamis, and Kami Vaniea. 2022. Virtual Reality Observations: Using Virtual Reality to Augment Lab-Based Shoulder Surfing Research. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 291–300. <https://doi.org/10.1109/VR51125.2022.00048>
  - [52] Robert Miller, Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Realtime Behavior-Based Continual Authentication of Users in Virtual Reality Environments. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 253–2531. <https://doi.org/10.1109/AIVR46125.2019.00058>
  - [53] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2021. Using Siamese Neural Networks to Perform Cross-System Behavioral Authentication in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 140–149. <https://doi.org/10.1109/VR50410.2021.00035>
  - [54] Tahrira Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. Unsure how to authenticate on your vr headset? come on, use your head!. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*. 23–30.
  - [55] Yoshitaka Oku. 2022. Temporal variations in the pattern of breathing: techniques, sources, and applications to translational sciences. *The Journal of Physiological Sciences : JPS* 72 (2022). <https://api.semanticscholar.org/CorpusID:251911297>
  - [56] Ilesanmi Olade, Hai-Ning Liang, Charles Fleming, and Christopher Champion. 2020. Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr). In *Proceedings of the 2020 4th international conference on virtual and augmented reality simulations*. 45–52.
  - [57] Marie-Caroline Play, Robin Trama, Guillaume Y Millet, Christophe Hautier, Marlene Giandolini, and Jérémy Rossi. 2022. Soft tissue vibrations in running: a narrative review. *Sports Medicine-Open* 8, 1 (2022), 131.
  - [58] JM Reina, José Manuel García-Aznar, J Dominguez, and M Doblaré. 2007. Numerical estimation of bone density and elastic constants distribution in a human mandible. *Journal of Biomechanics* 40, 4 (2007), 828–836.
  - [59] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1379–1384.
  - [60] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/>
  - [61] Yiran Shen, Hongkai Wen, Chengwen Luo, Weitao Xu, Tao Zhang, Wen Hu, and Daniela Rus. 2018. GaitLock: Protect virtual and augmented reality headsets using gait. *IEEE Transactions on Dependable and Secure Computing* 16, 3 (2018), 484–497.
  - [62] James M Wakeling, Benno M Nigg, and Antra I Rozitis. 2002. Muscle activity damps the soft tissue resonance that occurs in response to pulsed and continuous



- vibrations. *Journal of applied physiology* 93, 3 (2002), 1093–1103.
- [63] Ruxin Wang, Long Huang, and Chen Wang. 2023. Low-effort VR Headset User Authentication Using Head-reverberated Sounds with Replay Resistance. In *2023 IEEE Symposium on Security and Privacy (SP)*. 3450–3465. <https://doi.org/10.1109/SP46215.2023.10179367>
- [64] Mary A. Woo, William G. Stevenson, Debra K. Moser, Robert B. Trelease, and Ronald M. Harper. 1992. Patterns of beat-to-beat heart rate variability in advanced heart failure. *American heart journal* 123 3 (1992), 704–10. <https://pubmed.ncbi.nlm.nih.gov/1539521/>
- [65] 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>
- [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 *Financial Cryptography and Data Security: FC 2015 International Workshops, BITCOIN, WAHC, and Wearable, San Juan, Puerto Rico, January 30, 2015, Revised Selected Papers*. Springer, 281–297.
- [67] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. Glassgesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.
- [68] Zhen Yu, Hai-Ning Liang, Charles Fleming, and Ka Lok Man. 2016. An exploration of usable authentication mechanisms for virtual reality systems. In *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE, 458–460.
- [69] 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*. 446–459.
- [70] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous authentication using eye movement response of implicit visual stimuli. *proceedings of the acm on interactive, mobile, wearable and ubiquitous technologies* 1, 4 (2018), 1–22.
- [71] Huadi Zhu, Wenqiang Jin, Mingyan Xiao, Srinivasan Murali, and Ming Li. 2020. Blinkkey: A two-factor user authentication method for virtual reality devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–29.
- [72] Huadi Zhu, Mingyan Xiao, Demoria Sherman, and Ming Li. 2023. SoundLock: A Novel User Authentication Scheme for VR Devices Using Auditory-Pupillary Response. *Proceedings 2023 Network and Distributed System Security Symposium (2023)*. <https://api.semanticscholar.org/CorpusID:257501425>