HT-Auth: Secure VR Headset Authentication via Subtle Head Tremors

ZHIXIANG HE, School of Cyber Science and Engineering, Wuhan University, China
FENGYUAN RAN, School of Cyber Science and Engineering, Wuhan University, China
JING CHEN, School of Cyber Science and Engineering, Wuhan University, China
YANGYANG GU, School of Cyber Science and Engineering, Wuhan University, China
KUN HE, School of Cyber Science and Engineering, Wuhan University, China
RUIYING DU, School of Cyber Science and Engineering, Wuhan University, China
JU JIA, School of Cyber Science and Engineering, Southeast University, China
CONG WU, Department of Electrical and Electronic Engineering, University of Hong Kong, China

While Virtual Reality (VR) applications have gained popularity in recent years, efficiently identifying users on VR devices remains challenging. Current solutions, such as passwords and digital PINs, relying on handheld controllers or in-air hand gestures, are time-consuming and far less convenient than typing on touchscreens or physical keyboards. Even worse, the entry process can be observed by others in proximity, raising security concerns. In this paper, we propose HT-Auth, a novel authentication method for VR devices based on subtle head tremors. These tremors, occurring during active force exertion, are intrinsic and inevitable for human beings, which can be easily captured by inertial sensors built-in commodity VR headsets. We thus derive neck muscular biometrics from the tremor signal for secure VR device authentication. Our experiments, conducted with both standalone and mobile VR headsets, achieve a commendable balanced accuracy of 97.22% with just 10 registration samples, proving its efficacy and resilience against potential threats. Our source code is available at https://anonymous.4open.science/r/HT-OpenSource-10C3/.

CCS Concepts: • **Security and privacy**.

Additional Key Words and Phrases: Biometrics, User Authentication, Virtual Reality

ACM Reference Format:
Zhixiang He, Fengyuan Ran, Jing Chen, Yangyang Gu, Kun He, Ruiying Du, Ju Jia, and Cong Wu. 2025. HT-Auth: Secure VR Headset Authentication via Subtle Head Tremors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 85 (September 2025), 26 pages. https://doi.org/10.1145/3749980

Authors' Contact Information: Zhixiang He, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, zhixianghe@whu.edu.cn; Fengyuan Ran, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, rfy_Reflow@whu.edu.cn; Jing Chen, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, chenjing@whu.edu.cn; Yangyang Gu, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, guyangyang@whu.edu.cn; Kun He, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, hekun@whu.edu.cn; Ruiying Du, School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China, duraying@whu.edu.cn; Ju Jia, School of Cyber Science and Engineering, Southeast University, Nanjing, JiangSu, China, jiaju@seu.edu.cn; Cong Wu, Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong, China, congwu@hku.hk.

ACM 2474-9567/2025/9-ART85
https://doi.org/10.1145/3749980

1 INTRODUCTION

Motivation. Recently, virtual reality (VR) technology has gained widespread recognition among the public. The shipments of VR headsets are projected to reach 24.19 million in 2025, with an estimated compound annual growth rate of 17.36% [2]. Blending digital and physical worlds, VR is permeating a variety of fields, including entertainment, education, healthcare, and e-commerce to provide immersive experiences. This inevitably migrates vast sensitive data and functionalities from smartphones and computers to VR devices, including application accounts, browsing habits, and financial records. Therefore, a secure authentication system is crucial for VR devices to resist unauthorized access.

However, user authentication on VR devices is still in its early stages. Following conventional personal device-granting solutions, current methods rely on handheld controllers or in-air hand gestures for password entry, which are tedious and far less efficient than typing on touchscreens or physical keyboards. What's worse, the entry process could be observed by others in proximity and inferred credentials by mapping movement trajectories and the virtual keyboard layout [16, 19, 58, 77]. State-of-the-art research demonstrates that adversaries can successfully decipher 75% of text inputs just by observing victims' hand gestures from a distance [19].

There have been active works to explore alternative authentication solutions [10, 38, 40, 41, 45, 47, 53, 78], most of which involve time-consuming tasks in the virtual space or rely on additional hardware. For example, the user's head movements have been proven to effectively identify users when they listen to music beats [38], move the VR pointer to follow/throw a ball [45, 47], or walk [53]. These tasks take time, for instance, distinguishing users by head movements to music beats typically requires 5–10 seconds of listening [38]. Some studies utilize invariable body traits, such as electrooculogram (EOG) [41], electroencephalogram (EEG) [40], and electromyogram (EMG) [10], where dedicated and costly sensors are required. A recent study has explored the feasibility of using auditory-pupillary responses as biometrics [78], but it relies on an eye tracker that many commodity VR devices lack. Moreover, it requires a relatively long enrollment time (over 800s), making it inefficient to use.

Our approach. In this paper, we introduce a new kind of biometric indicator, head tremors (HT), for secure user authentication on VR headsets. To authenticate, a user simply needs to tense their neck skeletal muscle actively. Our approach, HT-Auth, is grounded in the key observation that head tremors, generated by users during neck skeletal muscle tensing, are intrinsic and inevitable for human beings. In this process, spinal cord neurons periodically send signals that stimulate neck muscle fibers to contract in a subtle, continuous cycle [12, 35]. The biological uniqueness of the neck muscular structures makes it possible to explore head tremors for user identification. During authentication, such tremors can be detected by commodity VR headsets that feature low-end built-in inertial sensors. A user is granted personal access if their login data matches the template created during the enrollment stage.

We encounter several challenges to realize HT-Auth: *(1) Presence of artifacts caused by human motions:* In virtual environments, users could engage in strong or micro movements while utilizing HT-Auth (e.g., head/body rotation, walking, speaking), which interfere with the precise detection and segmentation of head tremor signals. *(2) Unexplored biometrics related to head tremors:* The impact of neck muscle-related biometrics on head tremors has not been explored in prior work. How to extract unique and consistent muscular-associated biometric representations from the tremor signals is a challenge. *(3) Inconsistency in user behaviors during tremor generation:* The user's head tremor behaviors exhibit variations each time, resulting in different durations and densities of the tremor signals. How to mitigate the influence of this behavior inconsistency and guarantee reliable user authentication is a challenge.

To address the first challenge, we analyze the impact of different types of noises on head tremor signals. We then employ a set of signal processing techniques, such as power spectral density (PSD) analysis and GA-based MODWT denoising (Sec. 3.2), to enhance the tremor signal's SNR tailored to the specific noise characteristics. Moreover, we propose a specially designed three-stage event localization algorithm based on aggregated signal variance

(Sec. 3.3), which could adaptively detect the occurrence of head tremors. To tackle the second challenge, we investigate and validate the distinct biometric traits present in head tremors (Sec. 2.2, 2.4). From this foundation, we derive muscle physics-level features from two aspects: muscular contraction (Sec. 3.4.1) and muscular endurance (Sec. 3.4.2). To address the third challenge, we implement a lightweight feature reconstruction model based on a Siamese network to obtain tremor behavior-irrelevant features (Sec. 3.4.3). Coupled with transfer learning, this enables quick adaptation into a user-specific model with minimized registration samples (e.g., 10 samples), ensuring accurate authentication of newly registered users (Sec. 3.5). Our contributions can be summarized as follows:

- We propose HT-Auth, a secure authentication system that can be seamlessly integrated into off-the-shelf VR devices. It is the first work that shows distinctive neck muscular biometrics from head tremor signals can be extracted using built-in VR inertial sensors.
- We propose a specially designed three-stage algorithm for event localization, combining PSD analysis and GA-based MODWT denoising, which enables HT-Auth to reliably capture tremor patterns from redundant raw recordings and produce de-noised signal segments.
- We design a novel biometric representation that characterizes the unique contraction and endurance characteristics of human neck muscles. Capitalizing on the strength of the Siamese network and transfer learning, we mitigate the impact of behavioral inconsistencies on the features, enabling high-accuracy user authentication with minimal registration samples.
- We conduct extensive experiments to evaluate HT-Auth's effectiveness using a standalone and a mobile phone VR headset. The results demonstrate that it achieves a high balanced accuracy (BAC) of 97.22% with only 10 registration samples.

2 PRELIMINARY

2.1 Threat Model

The adversary's goal is to cheat HT-Auth for bypassing the authentication. We assume the adversary can physically access the victim's VR device when it is left unattended or stolen. We primarily consider the following attacks in this work. Other sophisticated attacks will be discussed in Sec. 7.

Blind attack. An adversary has no idea how HT-Auth works. So he simply puts on the victim's device and rotates the head arbitrarily, hoping the behavior can help bypass the authentication.

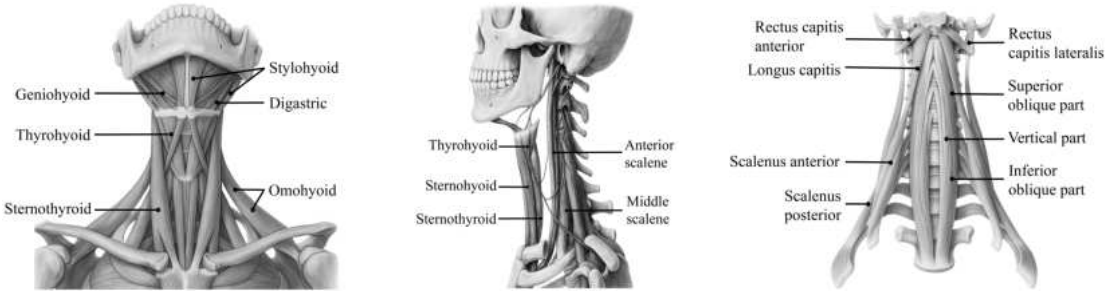
Impersonation attack. An adversary observes the victim's head tremors during authentication and attempts to replicate the behavior using their own neck, hoping to trick HT-Auth into granting access.

Synthesis attack. An adversary attempts to record the victim's head tremors using cameras, and then uses video processing techniques to synthesize the corresponding IMU signals. These signals are injected into the authentication system directly, bypassing the sensors in the process.

2.2 Kinetics of Head Tremor Production

In this section, we first outline the basic structure of human neck skeletal muscles, and then introduce how these muscles generate head tremors.

The basic structure of human neck skeletal muscles. A human neck skeletal muscles can be divided into three parts [17]: anterior muscles, lateral muscles, and posterior muscles, as shown in Fig. 1. These skeletal muscles are typically symmetrically distributed on both sides of the cervical spine, serving to protect the cervical spine, facilitate head movement, and support head stability. The functional unit of skeletal muscle is the motor unit [48], comprising a single motor neuron and all the muscle fibers it innervates. Muscle fibers within a motor unit typically belong to the same type, i.e., type I and II. The former has a relatively smaller diameter, contracts slowly, and exhibits greater endurance; the latter has a larger diameter, and contracts quickly to generate explosive



(a) Anterior muscles (b) Lateral muscles (c) Posterior muscles

Fig. 1. Human neck muscles from different perspectives.

strength, but fatigues rapidly. A skeletal muscle contains multiple motor units, and when a motor neuron is stimulated, all muscle fibers within that unit contract simultaneously.

The rationale for head tremors. Head tremors are subtle physiological tremors characterized by low-amplitude, rapid back-and-forth movements of the head, typically occurring below 15Hz [35, 56]. These tremors arise from isometric contractions of the neck skeletal muscles [48], where muscle fibers generate tension to hold the head in a fixed position without joint movement. To sustain this tension, spinal cord neurons periodically send signals that trigger muscle fibers to contract, a mechanism known as the stretch-reflex circuitry [12, 35]. This rhythmic activation mechanism causes the muscle fibers to contract and relax in a subtle, continuous cycle, resulting in head tremors. Due to differences between physical structures of neck skeletal muscles among individuals (e.g., the number of motor units, the size/shape of muscle fibers, the ratio of fiber type, etc.), the frequency responses of head tremor signals vary greatly from person to person. Building on the theoretical analysis, we next conduct experiments to explore the feasibility of utilizing head tremor signals for user authentication.

2.3 Head Tremor Tracking in VR

VR headsets determine their position and orientation in the real world with the Inertial Measurement Unit (IMU) sensor. The IMU, comprising an accelerometer and a gyroscope, tracks six Degrees of Freedom (DoF). Specifically, the accelerometer measures linear acceleration across the x, y, and z axes in m/s^2, while the gyroscope measures angular velocity along the same axes in rad/s. These raw readings are sent to the VR software, which adjusts the display accordingly. By integrating these raw measures, the software outputs the headset's position (denoted as P_x, P_y, P_z) and orientation (denoted by a quaternion: O_x, O_y, O_z, O_w), totaling seven axes, which can be accessed through an interface, enabling the tracking of head tremors.

2.4 Feasibility Study

To study the feasibility of using head tremor signals for authentication, we recruit 4 subjects to perform head tremors at rest, with each contributing 50 samples. Fig. 2 presents the Power Spectral Density (PSD) profiles of the signals (O_z) collected from four subjects, which are notably concentrated in the low-frequency range. To get a clear picture of differences among subjects, we calculate the average Person Correlation Coefficients (PCCs) of PSDs between subjects 1 and 4, subjects 2 and 4, and subjects 3 and 4, which are 0.55, 0.57, and 0.51, respectively. While the average PCCs between different trials for subject 4 is 0.82. We further use the t-distributed Stochastic Neighbor Embedding (t-SNE) projections [61] to visualize the PSD patterns for all trials. As shown in Fig. 3, different clusters exhibit distinct centroids, suggesting head tremors could be used for user distinguishing.

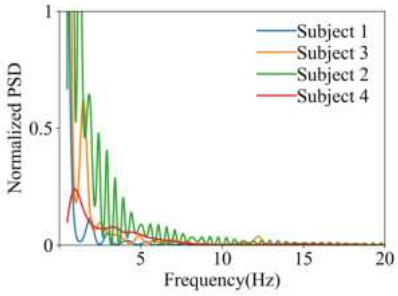


Fig. 2. PSD profiles of head tremor signals from 4 subjects.

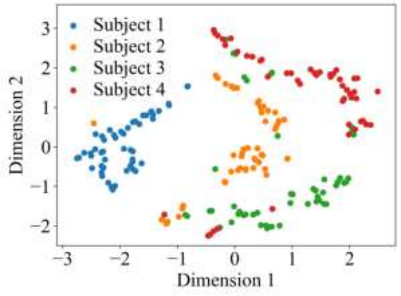


Fig. 3. t-SNE visualization of signal PSDs from 4 subjects.

However, the boundaries between clusters are not well-defined, necessitating additional design to ensure a reliable authentication. In the next section, we explore some solutions for tackling this problem.

2.5 Exploring Behavioral Inconsistency Mitigation

In Sec. 2.4, the blurring of cluster boundaries stems from the inconsistency in subjects' head tremor behaviors. Fig. 4 illustrates the time-domain head tremor signals (O_z) across different trials from a single subject, revealing significant variations in duration and amplitude. This variability undermines the system's authentication performance, increasing the likelihood of false rejections. We explore several methods to address this issue, but none proved to be effective. We present them below and summarize the key limitations of each.

Re-sampling. The simplest approach is to normalize each time series into the same length via up-sampling (e.g., nearest neighbor interpolation [51]) or down-sampling (e.g., decimation factor [34]) algorithms. However, this process can lead to information loss and distort the signal's frequency characteristics. For instance, signals originally sampled at 500Hz can change significantly after re-sampling. Fig. 5 depicts the PSDs' t-SNE visualization after signal re-sampling, with the length of each signal standardized to 2000 data points, where the cluster boundaries remain blurred.

Dynamic Time Warping. Dynamic Time Warping (DTW) [46] finds the optimal alignment between two time series x_1 and x_2, allowing x_2 to be stretched or compressed to match x_1. DTW is effective for aligning signals with varying pacing or tempo, such as gait and speech, which may not be suitable for head tremor signals. Fig. 6 shows that DTW alignment does not enhance the clustering results.

Training Behavior-specific Models. Some studies train separate authentication models for each behavior [55]. When a user logs in, their behavior pattern is identified first, and then the corresponding authentication model is activated. However, this approach not only increases computational overhead but also adds user registration burdens, as each model requires a separate set of training samples.

Overall, none of these methods are ideal solutions for mitigating behavioral inconsistency in head tremors. In the following section, we attempt to tackle this problem through fine-grained feature extraction and reconstruction, which prove effective for head tremor signals.

3 SYSTEM DESIGN

In this section, we introduce the system architecture and design details for HT-Auth.

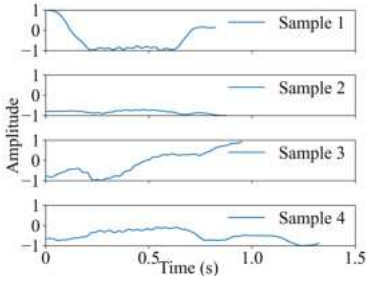


Fig. 4. Different samples from one subject.

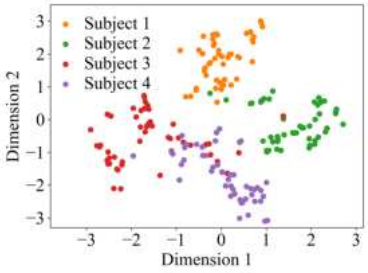


Fig. 5. t-SNE visualization of sample PSDs after processing by re-sampling.

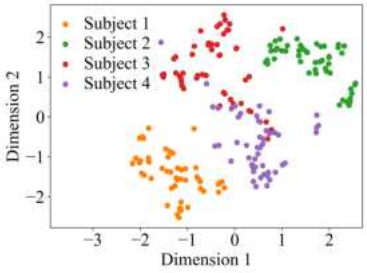


Fig. 6. t-SNE visualization of sample PSDs after processing by DWT.

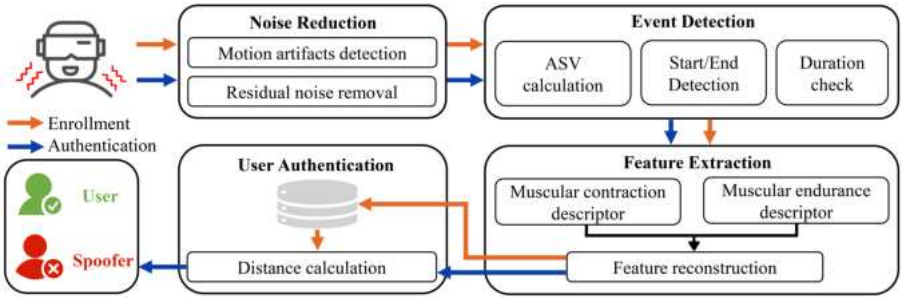


Fig. 7. Workflow of HT-Auth.

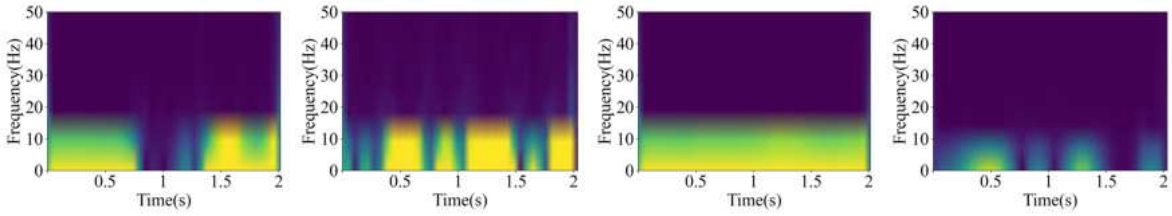
3.1 System Overview

HT-Auth operates in two phases: enrollment and authentication. During enrollment, the system creates users' profiles with data captured by the inertial sensors in the VR headset. During authentication, these profiles serve as templates for matching incoming sensor data.

Fig. 7 presents the four primary modules of HT-Auth: (1) Noise Reduction, (2) Event Detection, (3) Feature Extraction, and (4) User Authentication. Human macro motions (e.g., head/body rotation, walking) and micro movements (e.g., facial expression change, breathing) can alter sensor readings and pollute the head tremor signals. Thus, we employ a series of techniques, including PSD detection and GA-based MODWT, to denoise the signal. We then design a three-stage localization algorithm to isolate the head tremor signal with varying duration and amplitude. After that, muscle-physics level features are derived based on the head tremor rationale analysis. A Siamese network is designed to reconstruct features for alleviating the impact of human behavior inconsistency. Finally, these reconstructed features are fed into the authentication module for user distinguishing.

3.2 Noise Reduction

3.2.1 Motion Artifacts Removal. Human motions (e.g., head/body rotation, walking) have a frequency range of 0-20Hz [30], overlapping with head tremors that span 0-15Hz. As shown in Fig. 8, we find head tremor signals typically exhibit lower power than the other motion-induced signals, as indicated by the brightness in the spectrograms. To distinguish them, we apply a sliding window of width L across the signal and compute the PSD for each frame. Frames with PSD values exceeding a threshold are considered to be contaminated by human motion and are discarded.



(a) Head rotation (b) Body rotation (c) Walking (d) Head tremor

Fig. 8. The spectrograms of motion artifact signals (a,b,c) and head tremors (d).

3.2.2 Residual Noise Removal. The residual noises primarily stem from micro movements, such as facial muscle contraction during expression change (0.1-10Hz [64]) and slight body displacements during breathing (0.16-0.6Hz [9, 73]). The spectral subtraction technique [15, 37], traditionally used for speech enhancement, is unsuitable for these noises as they share frequency bands with the target signal, which may result in the incorrect subtraction of non-noise components. In our study, we denoise head tremor signals with the Maximal Overlap Discrete Wavelet Transform (MODWT) technique [13]. Compared with Short-Time Fourier Transform (STFT), MODWT excels at handling non-stationary signals (e.g., head tremors) due to the wavelet's excellent localization properties.

MODWT is a linear filtering operation that transforms a discrete signal into coefficients across various time and frequencies. These coefficients can be calculated with the well-known pyramidal algorithm [42]. The main idea is to decompose the signal into high- and low-frequency components, with the low-frequency part recursively decomposed into coefficients at finer frequency bands. Specifically, given a discrete signal $\{X_t\}, t = 0, \ldots, N-1$, a mother wavelet is selected to derive a pair of filters: a high-pass filter h and a low-pass filter g. MODWT performs multi-level filtering using these filters in a pyramidal fashion. At each level j, the decomposition is performed according to formulas $W_j = h * V_{j-1}$ and $V_j = g * V_{j-1}$, where $j = 1, \ldots, J$, $V_0 = \{X_t\}$, and $*$ represents the convolution operation. This yields a set of coefficients $\{W_1, W_2, \ldots, W_J, V_J\} \in \mathbb{R}^{\{J+1\}\times N}$ as the output of pyramidal algorithm. As the filters h and g are designed to halve the input frequency at each level, the resulting frequency bands of these coefficients are $W_j \in [fs/2^j, fs/2^{j-1}]$, $j = 1, \ldots, J$ and $V_J \in [0, fs/2^J]$, where f_s is the sampling frequency of $\{X_t\}$. These bands are non-uniformly distributed, offering finer resolution at lower frequencies, thus well-suited for analyzing head tremor signals.

MODWT disentangles the signal, allowing head tremors to manifest as strong coefficients, while residual noise, such as expression change or breathing, dissolves into weaker ones. The basic idea of MODWT-based denoising is to diminish these weaker coefficients, allowing the reconstructed signal to be cleaner and primarily retain the tremor components. The denoising performance relies on the MODWT parameter configuration, including the mother wavelet ψ, decomposition level δ, signal extension rule ξ, and thresholding type τ. Mother wavelet ψ is the core of MODWT, defining the signal's local characteristics and guiding how it is decomposed into coefficients across different time and frequencies. Decomposition level δ sets the resolution of these coefficients. A level that is too low may overlook important signal details, while a level that is too high can lead to overfitting. Signal extension rule ξ governs how the signal's boundary effects are handled, with a poor choice potentially causing distortion in the decomposed signal. Finally, thresholding type τ determines how thresholds are applied during denoising, balancing noise suppression with the preservation of key signal characteristics. Inspired by [3], we use a Genetic Algorithm (GA) to automatically select the optimal denoising parameters rather than determining them manually. Specifically, we define the genetic code of individual i as $G_i = \{\psi_i, \delta_i, \xi_i, \tau_i\}$, with each G_i randomly

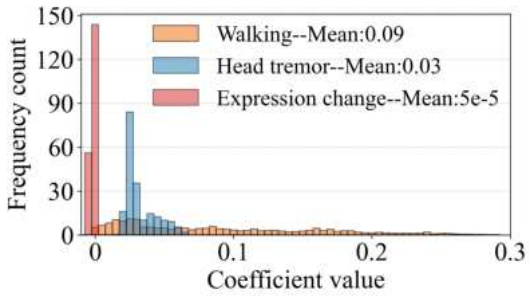


Fig. 9. Frequency count of MODWT-decomposed coefficients.

initialized in the population. The fitness function is formulated as the Minimum Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(x(n) - \hat{x}(n))^2$$
$$\hat{x}(n) = MODWT(G_i, x(n)) \tag{1}$$

Here, $x(n)$, $\hat{x}(n)$, and N represent the original signal, the denoised signal, and the total number of samples, respectively. Guided by the fitness function, we iteratively optimize parameter set G. We define the termination criteria as the population's fitness shows little change over successive generations. Once satisfied, the parameter set G with the best fitness is selected as the optimal denoising parameter set. With these parameters, we decompose, threshold, and reconstruct the data to produce the denoised head tremor signals.

Question: why cannot MODWT alone handle two noise types (i.e., motion artifacts and residual noise) simultaneously? To illustrate this issue, we analyze the MODWT decomposition coefficients of three types of signals: walking (motion artifacts), head tremor, and facial expression changes (residual noise). Fig. 9 presents the frequency count of the coefficients (V_J), as these signals primarily occupy the low-frequency bands. We observe a significant overlap between the decomposed coefficients of walking and head tremor, making it difficult to remove walking noise through coefficient filtering. We speculate that this is because walking is an intense and multi-joint coordinated movement, resulting in a wider range of decomposition coefficients. In contrast, facial expression coefficients are smaller and do not overlap with those of head tremors. By suppressing these coefficients, we can effectively remove the residual noise.

3.3 Event Detection

When a head tremor occurs, the inertial sensor captures data from various axes. Traditional methods typically rely on a reference axis for event detection [8, 37, 72]. However, due to significant behavioral variability across trials, the signal amplitude from the same axis can fluctuate greatly, i.e., weak at times and strong at others (see Fig. 4), rendering them unreliable for event detection. To accurately isolate head tremor signals and exclude non-tremor components, we develop a head tremor localization algorithm based on Aggregated Signal Variance (ASV), which consists of three steps applied to the entire activity signal:

Step1: ASV calculation. We apply a sliding window of width L on each axis. Let x_i^a and μ_i^a represent the head tremor signal and its mean on axis a within the i-th window. The ASV for the i-th window is calculated as:

$$ASV_i = \sum_{a=1}^{7}\frac{1}{L}\sum_{m=1}^{L}\left(x_i^a[m] - \mu_i^a\right)^2 \tag{2}$$

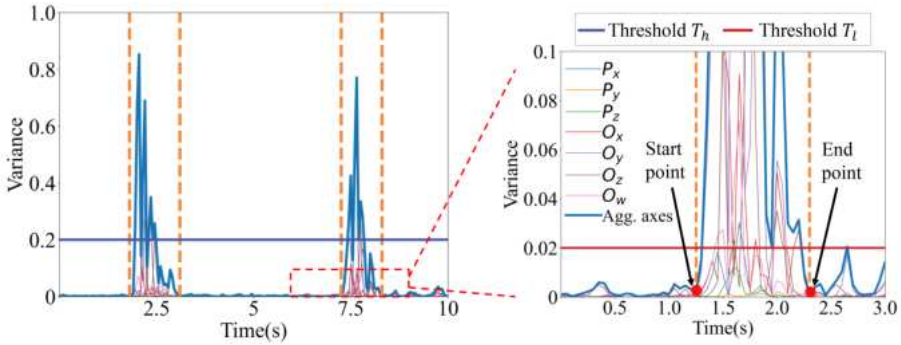


Fig. 10. The process of event detection.

Fig. 10 displays the single-axis variances and ASVs for two signals, where the ASVs are more prominent compared to single-axis variances.

Step2: Start/End detection. We modify a dual-threshold method [62] to identify the start and end points of each event. Specifically, we first set a threshold T_h on the ASV and find segments exceeding T_h, which are considered to contain events. Starting from each peak of these segments, we then traverse along the slopes on both sides to locate the start and end points of the events, defined as the first valley below $T_l (T_l < T_h)$. In contrast to the original method that determines event boundaries with the intersection of T_l and the signal, our optimized method better identifies entire events and reduces reliance on precise T_l settings.

Step3: Duration check. Once potential head tremor events are identified, we verify the duration of these activity segments. The activity is considered as a real head tremor only if its duration exceeds T_{dur}. Through extensive testing, we adjust the values of T_h, T_l, and T_{dur} to accurately locate head tremor events. Fig. 10 displays the detection results, where head tremor fragments are successfully segmented out.

3.4 Muscle Physics-level Feature Extraction

Our goal is to develop a set of person-distinguishable representations from head tremor signals. As described in Sec. 2.2, head tremors result from the subtle, continuous contraction and relaxation of neck muscles, we thus explore fine-grained representations that capture the uniqueness of these muscles.

3.4.1 Level-I: Muscular Contraction Descriptor. When neck muscles contract, variation in contraction speed and strength can influence the head tremor frequency and amplitude. For instance, motor units with predominantly type II muscle fibers produce faster, stronger contractions [48], leading to higher tremor frequencies and greater amplitudes. Thus, the frequency response of head tremor signals characterizes an individual's unique neck muscle contractions, termed the muscular contraction descriptor.

We use Mel Frequency Cepstral Coefficients (MFCC), Spectral Centroid (SC), and Spectral Spread (SS) to describe the frequency response of head tremors. MFCC, customized with Mel filter banks, is more effective than STFT for capturing key patterns in head tremor signals, which are dominated by low frequencies. The spectral centroid represents the spectrum's center of mass, while the spectral spread measures its dispersion. We segment the data from each axis into k frames, extracting MFCC, SC, and SS features per frame. For MFCC, we calculate k coefficients and average them across frames to describe the overall spectral profile. This yields features $\{F_{MFCC}(1), \ldots, F_{MFCC}(k), F_{SC}(1), \ldots, F_{SC}(k), F_{SS}(1), \ldots, F_{SS}(k)\}$ per axis, and the muscular contraction descriptor is assembled by aggregating these features across all seven axes.

3.4.2 Level-II: Muscular Endurance Descriptor. In the average population, about 50%-55% of muscle fibers are type I, and 45%-50% are type II, with these percentages being genetically determined and vary greatly among

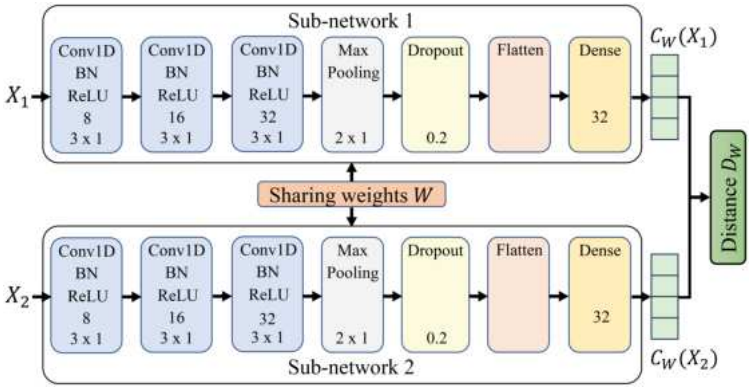


Fig. 11. Siamese network architecture.

individuals [14, 18, 48]. Type I fibers have greater endurance, so individuals with a higher proportion are more likely to have steadier head tremors. Thus, the signal's temporal consistency can reflect the unique endurance traits of an individual's neck muscles, termed the muscular endurance descriptor.

We use Spectral Flux (SF), Spectral Entropy (SE), and Zero-Cross Rate (ZCR) to describe the temporary consistency of head tremor signals. Spectral flux tracks how the spectrum shifts between frames, providing insight into the signal's dynamic variations over time. Spectral entropy measures the signal randomness or unpredictability. As muscular fatigue accumulates during activity, weakened control over contractions tends to make head tremors more unpredictable, thereby impacting spectral entropy. The zero-crossing rate measures how often a signal flips its sign in the time domain, helping to identify varying states in head tremors. Similarly, we segment data from each axis into k frames and extract features $\{F_{SF}(1), \ldots, F_{SF}(k), F_{SE}(1), \ldots, F_{SE}(k), F_{ZCR}(1), \ldots, F_{ZCR}(k)\}$ for each axis. We incorporate the spectrum shifts between the start and end frames into the SF features for dimension consistency. The muscular endurance descriptor is compiled by combining these features across all seven axes. As neck muscles typically adapt to the headset's weight and position in the short term and naturally evolve over the long term, we explore how the muscle's mid-term adaptation and long-term evolution influence the authentication performance, which could be seen in Sec. 5.2.9 for detail.

3.4.3 Behavioral Inconsistency Mitigation through Feature Reconstruction. In Sec. 2.5, we attempt to address behavioral inconsistency by adjusting the original signal, but the results fell short. The fundamental reason is that all these methods fail to remove the variable behavioral traits embedded in the signal. In this section, we resort to deep learning, i.e., a Siamese network, for feature reconstruction. Through contrastive learning, the network learns to filter out variable behavioral traits, retraining stable and highly recognizable physical biometrics in the reconstructed representations.

Base model structure. As depicted in Fig. 11, the idea of the Siamese network is basically to utilize twin sub-networks with identical architecture and shared weights to compare the similarity of two inputs [11, 33]. Each sub-network begins with a convolutional layer with 8 kernels of size 3×1 to learn larger scale features, followed by two convolution layers with 16 and 32 kernels of size 3×1 to learn small-scale features. Batch normalization and ReLU activation functions are applied after each convolutional layer. We also include a max-pooling layer with a kernel size 2×1 and a dropout layer with a dropout probability of 0.2 to prevent over-fitting. Finally, a flatten layer and a dense layer map the features into a fixed-size representation. Given a pair of feature vectors, the twin sub-networks output two representations, $C_W(X)$, and calculate their distance, D_W, to quantify similarity.

During training, we select feature vector pairs $\{X_a, X_p\}$ and $\{X_a, X_n\}$ as inputs to the Siamese network. The positive input X_p, comes from the same user as the anchor X_a, while the negative input X_n, belongs to a different

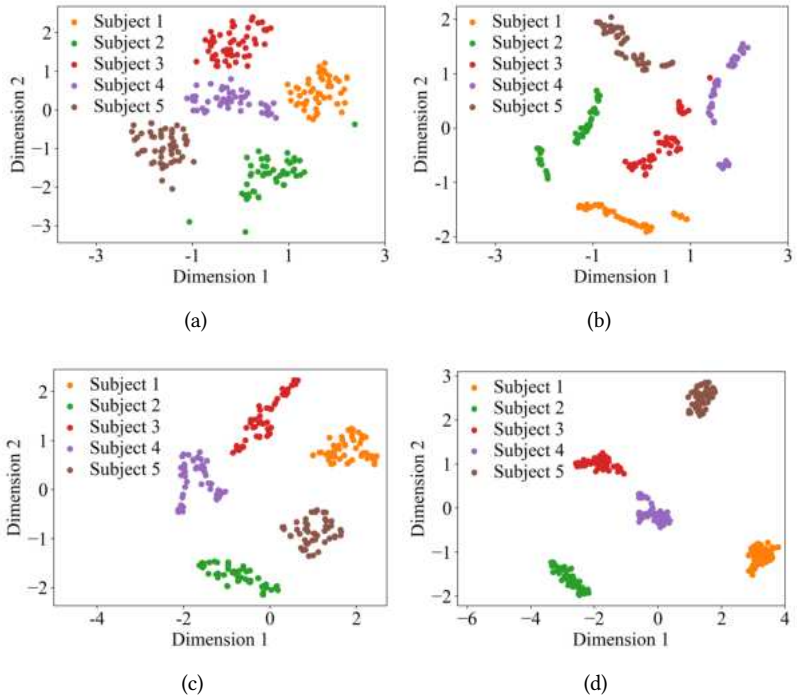


Fig. 12. t-SNE visualization of muscular contraction descriptor (a), muscular endurance descriptor (b), hybrid features (c), and reconstructed features (d).

user. The network's objective is to minimize the distance between X_a and X_p while maximizing the distance between X_a and X_n. To achieve this, a loss function $L(W)$ is formulated to optimize the parameters W of each sub-network.

$$L(W) = \sum_{i=1}^{N} Y(D_W^i)^2 + (1-Y)max(M - D_W^i, 0)^2 \tag{3}$$

Here, Y indicates whether the two inputs are from the same user, i.e, $Y = 1$ for X_a and X_p, and $Y = 0$ for X_a and X_n. N is the batch size and M is the margin representing the decrease interval.

Transfer the base model to a specific user. Transfer learning enables the base model to adapt to new user data without undergoing full retraining, leveraging its existing knowledge to minimize the number of registration samples needed from a new user. Specifically, once enough user samples are registered, we create feature pairs $\{X_a, X_p\}$ and $\{X_a, X_n\}$. X_a and X_p are randomly selected from the registered user set, while X_n are selected from the default user set. We assume that the default user set, solely used for model generation, has been preloaded into the VR headset upon shipment. These randomly selected pairs then serve as inputs for the fine-tuning process.

Feature visualization. We utilize t-SNE to visualize the feature space of the muscular contraction descriptor, muscular endurance descriptor, hybrid features (a combination of both descriptors), and the reconstructed features, as illustrated in Fig. 12. Overall, both descriptors demonstrate user distinctions, and combining them clarifies the cluster boundaries. After feature reconstructing, with behavior inconsistency mitigated, the samples exhibit smaller intra-class distances and larger inter-class distances. This validates the effectiveness of the extracted features for analyzing head tremors.

3.5 User Authentication

User authentication begins by extracting features from the login sample $\widetilde{X}$ and generating an embedding vector $C_W(X)$ with the Siamese sub-network. This vector is then compared to a pre-computed template, R, the average of the legitimate user's registered samples:

$$R = \frac{1}{N}\sum_{i=1}^{N} C_W(X_i) \tag{4}$$

where R is the template, $C_W(X_i)$ is the reconstructed feature, and N is the total number of registered samples. The login user is authenticated as legitimate if the Euclidean distance between $C_W(\widetilde{X})$ and R falls below a threshold.

4 IMPLEMENTATION

4.1 Experimental Setup

4.1.1 Design Details. For motion artifacts detection (Sec. 3.2.1), the choice of window length L affects denoising performance. A too-short window may fail to remove motion artifacts, while a too-long one may mistakenly remove head tremors. We set $L = 25$ as a trade-off based on experimental observation. The PSD values are summed across all seven axes for each frame, and if the result exceeds two times the typical head tremor value (obtained with statistical analysis), we consider it to be non-tremor motion. For residual noise removal (Sec. 3.2.2), the final parameters are set $\{\psi, \delta, \xi, \tau\} = \{db4, 4, per, soft\}$, which are determined through GA-based optimization. $db4$, or Daubechies 4 mother wavelet, is known for its excellent localization properties, making it well-suited for describing transient signals like head tremors. per, or periodic boundary extension, reduces edge effects by assuming the signal repeats at its boundaries. $soft$, or soft thresholding, is a denoising technique that zeros out smaller coefficients and gently reduces larger ones, helping to suppress residual noise while maintaining a smoother reconstructed signal. For the event detection (Sec. 3.3), we again adopt a window size of $L = 25$. The parameters T_h, T_l, and T_{dur} affect the performance of event detection. Specifically, a high T_h may overlook valid tremors, while a low one may mistake weak noise as tremors. Improper T_l (too high or too low) may lead to inaccurate detection of event start and end points. T_{dur} that is too long may discard valid tremors, while one that is too short may mistake short-duration noises as tremors. After calculating the ASV, we normalize the amplitude and set $T_h = 0.2$, $T_l = 0.02$, and $T_{dur} = 0.25s$ for segmentation, guided by experimental observation. For feature extraction (Sec. 3.4), we set $k = 13$, aligning with the feature dimension utilized in many existing studies [23, 24]. The hybrid feature dimension of the two descriptors is $6 \times 7 \times 13 = 546$, which is fed into the feature reconstruction model, producing $C_W(X)$ with a dimensionality of 32, a value commonly adopted as embedding dimension in prior researches [25, 26]. In user authentication (Sec. 3.5), a predefined threshold is required for distinguishing users. In our implementation, the threshold is a fixed value (available in the open-source code) for all subjects and determined through experimental observation. We use grid search to find the optimal threshold that maximizes BAC on the pre-training set (see Sec. 4.1.2). Since the pre-training data is collected from the authentication scenario, this threshold can also be used to distinguish the legitimacy of other users.

4.1.2 Siamese Network Training. The Siamese network's training process can be divided into two stages: pre-training and transfer learning. Initially, we randomly initialize the network parameters and pre-train the network with data from 10 subjects (i.e., regarded as default users), each contributing 50 samples. The data collection details are available in Sec. 4.2. To get network inputs, we alternate each default user as the legitimate user, extracting features from their samples (anchor input X_a), pairing them with either their features (positive input X_p) or those from other users (negative input X_n). We fine-tune hyperparameters with grid search and optimize the network using the Adam optimizer [32]. The network is pre-trained only once and can be adapted to different

(a) (b) (c) (d)

Fig. 13. The devices employed: Meta Quest 2 (a), Shinecon with IQOO Neo6 SE (b); user-wearing scenarios: wearing Meta Quest 2 (c), wearing Shinecon (d).

users through transfer learning. During transfer learning, we select X_a and X_p from the user's enrolled data, and X_n from the dataset of default users to create a personalized profile.

4.1.3 HT-Auth Prototype. We evaluate our system with two types of VR headset devices: a standalone VR (i.e., Meta Quest 2) and a mobile phone VR (i.e., Shinecon VR with a smartphone - IQOO Neo6 SE). The devices employed and the user-wearing scenario are shown in Fig. 13. We build two experimental platforms for this study. For the standalone VR, we develop a VR App based on OpenVR SDK v2.5.1, utilizing the built-in *getRawTrackedDevicePoses* function to get the headset's 3-axis position and 4-axis orientation data. For the mobile phone VR, we build an Android App and install it on a smartphone running Android 9.0. This App leverages the $getDefaultSensor$ function to get data from the phone's inertial sensors, which are later mathematically transformed into position and orientation data for consistency. All data is collected at $500Hz$ and processed offline.

4.2 Data Collection

We have recruited 30 subjects (8 females and 22 males) aged from 18 to 27 to conduct experiments with both VR devices. The subjects, comprise graduate and undergraduate students, with each given $ 5.5 as an incentive for participating. The IRB approval is obtained.

Before data collection, each subject is given time to use the VR headset and adjust the headset straps for comfortable tightness. We divide these subjects into two groups, i.e., *group-A* and *group-B*. Data from *group-A* (10 subjects) is used for Siamese network pre-training, while data from *group-B* (20 subjects) is used for system evaluation. Each subject participates in five sessions, repositioning the VR headset between each. In each session, subjects are asked to perform 10 head tremors, with a 6-second interval between each. This generate *dataset-1* with $10 \times 5 \times 10 = 500$ samples and *dataset-2* with $20 \times 5 \times 10 = 1000$ samples.

We also collect data across different scenarios, such as head and body gestures, device strap tightness, long-term effects, etc. For each specific condition (e.g., standing), data from five subjects are gathered in two sessions. We fine-turn each user-specific model with data from *dataset-2* and conduct testing under different scenarios.

4.3 Evaluation Metrics

The following metrics are utilized for HT-Auth evaluation. False Acceptance Rate (FAR): the ratio of illegal users who gain access. False Rejection Rate (FRR): the ratio of legitimate users who are denied access. Balanced Accuracy (BAC): the average between the True Acceptance Rate (TAR) [21, 22] and the True Rejection Rate (TRR) [20], which are commonly used for evaluating models trained from unbalanced data [6]. Frequency Count of Scores (FCS) [59]: the frequency count of samples' prediction scores. In HT-Auth, we define the prediction score as the opposite of distance D_W. We adopt these metrics collectively to evaluate the authentication system. Specifically, BAC provides a high-level view of overall system performance by measuring how well genuine and imposter samples are distinguished, and is therefore widely used in authentication studies [26, 70, 71]. However, BAC does

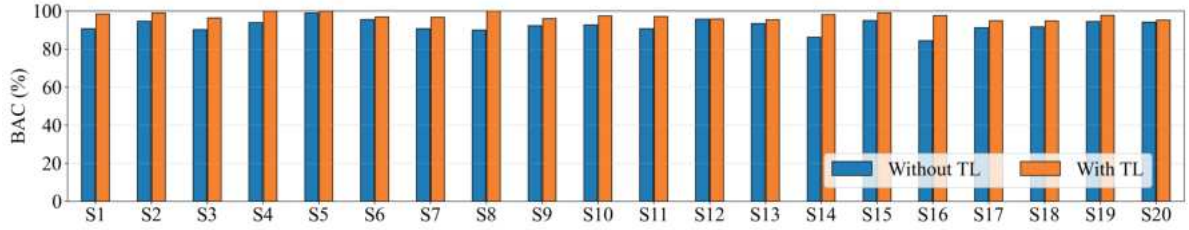


Fig. 14. Performance comparison of without/with transfer learning.

not provide separate discrimination information for genuine and imposter samples, which are complemented by FRR and FAR. Moreover, in real-world authentications, FRR is less dramatic than FAR, especially when a few immediate retries succeed, a point we evaluate in Sec. 5.2.9.

5 PERFORMANCE EVALUATION

5.1 Overall Performance

We evaluate the system's overall performance using *dataset-2*. To demonstrate HT-Auth's effectiveness in handling few-shot scenarios, we enroll each user with only 10 samples. The remaining legitimate samples, along with unauthorized ones from other users, are used for testing.

5.1.1 Effective of Transfer Learning. We evaluate transfer learning (TL) in two scenarios: i) without TL, get reconstructed features with the pre-trained Siamese network directly, and ii) with TL, get reconstructed features with user-specific models. During TL, 10 samples from the enrolled user are utilized to update the pre-trained network's parameters. Fig. 14 displays the BACs of 20 subjects under two scenarios. Without TL, the average BAC is 92.25%, which increases to 97.22% after applying TL. We also observe a performance gain for each subject with the application of TL, with subject 16 achieving the largest BAC increase of 13%. These improvements demonstrate the effectiveness of transfer learning, enabling the model to better adapt to individual user variations.

5.1.2 Performance of Different Feature Sets. We compare the performance of different feature sets, including Muscular Contraction Descriptor (MCD), Muscular Endurance Descriptor (MED), Hybrid Features (HF), and Reconstructed Features (RF). User templates are generated from each feature set, and the same authenticator in Sec. 3.5 is used for legitimacy verification. As shown in Fig. 15(a), the BACs for MCD, MED, HF, and RF are 87.78%, 85.53%, 91.03%, and 97.18%, respectively. We observe that although both descriptors demonstrate potential in user differentiation, they falter with certain individuals. Combining them together tackles this issue and yields additional performance improvements. Nonetheless, due to inherent inconsistencies in user behavior, the original hybrid features achieve only an average BAC of 91.03%. Implementing feature reconstruction mitigates these inconsistencies, resulting in a significant BAC enhancement of 6.15%.

5.2 Impact of Various Factors

5.2.1 Different Devices. We compare the performance of the standalone VR and the mobile phone VR, which represent two categories of popular VR models. Table 1 displays the authentication performance achieved by two VR models. Specifically, without TL, the BACs are 92.25% for standalone VR and 91.70% for mobile phone VR. Incorporating TL further enhances the performance, raising the BACs to 97.22% and 97.50%, respectively. Overall, our system demonstrates good performance on both VR models, suggesting its scalability and adaptability across different VR devices.

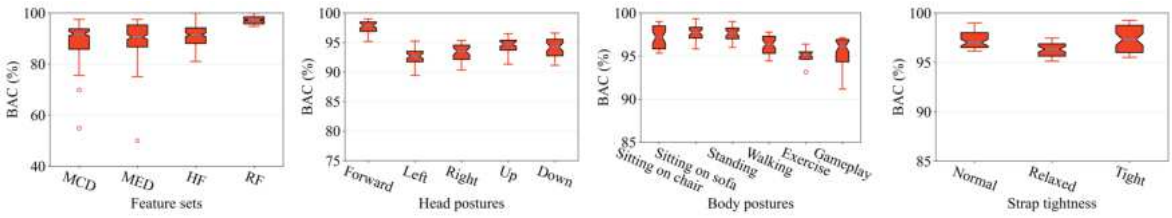
Table 1. Mean/standard deviation of BAC (%), FAR (%), and FRR (%) of two VR models.

	Channel	BAC	FAR	FRR
Without TL	Standalone VR	92.25/3.24	12.24/4.12	3.41/5.49
	Mobile phone VR	91.70/4.87	4.60/5.73	9.45/5.68
With TL	Standalone VR	97.22/1.68	1.41/3.36	3.85/2.30
	Mobile phone VR	97.50/3.93	1.32/1.30	4.16/3.48

5.2.2 Head and Body Postures. We evaluate the system performance under five head postures, i.e., facing forward, left, right, up, and down, with facing forward as the baseline. We fine-tune each user-specific model with 10 samples of facing forward, and test it under different postures. Fig. 15(b) depicts the BACs of five head postures, which are 97.47%, 92.53%, 93.17%, 94.33%, and 94.05%, respectively. Our analysis indicates that varying head postures can impact the authentication performance, leading to a BAC reduction ranging from 3% to 5%. To enhance generalization, we incorporate 5 samples from each of the other four postures (left, right, up, and down) into the training set, resulting in finally BACs of 97.36%, 96.93%, 97.13%, 96.87%, and 97.29% for facing forward, left, right, up, and down, respectively. We also consider six body postures, including static postures, i.e., sitting on chair, sitting on sofa, and standing; dynamic postures, i.e., walking, exercise (e.g., body stretching and torso twists), and interactive gameplay (e.g., beat saber). For dynamic postures, head tremors are performed during intervals between body movements. HT-Auth needs to accurately remove noise, identify tremors, and then perform authentication. Fig. 15(c) illustrates the BACs of these postures, which are 97.16%, 97.66%, 97.58%, 96.23%, 94.96%, and 95.13%, respectively. Overall, the system performs better in static than dynamic postures. When the motion artifact removal step (Sec. 3.2.1) is skipped, BAC drops to 82.74%, 85.29%, and 78.23% for walking, exercise, and interactive gameplay, given more noisy samples being mistakenly treated as genuine head tremors. We further assess the system performance when only using one-axis data (e.g., O_z) for event detection. The resulting BACs are 92.47%, 94.56%, 93.29%, 90.62%, 89.65%, and 90.36% for six body postures, showing a decrease of 3.1%–5.61% compared to ASV-based event detection. This is mainly because single-axis data is insufficient to accurately capture tremor events, leading to more genuine samples being rejected.

5.2.3 Template Transferability. We explore cross-device authentication, where a user enrolls on one device and uses the resulting biometric template R to match login samples on another. We consider two scenarios: i) enrolls on standalone VR and logs in on mobile phone VR; ii) enrolls on mobile phone VR and logs in on standalone VR. During enrollment, 10 samples from a legitimate user are used to create the template. Overall, the BACs for scenarios i) and ii) are 76.17% and 82.36%, respectively. This suggests that a template registered on one device can not be directly applied to another, as differences in shape, sensor placement, and strap design affect headset-head coupling and thus impact tremor signals. A simple strategy is to update the template using samples from the new device. For example, by updating the Siamese network parameters (transfer learning) and the template with 10 samples, the BACs for scenarios i) and ii) can reach 97.12% and 97.58%, respectively.

5.2.4 Device Strap Tightness. The strap tightness of the VR headset affects the coupling state between the device and the user's head, potentially influencing the authentication performance. We consider three levels of strap tightness: relaxed, normal, and tight. We distinguish them based on user perception: relaxed, the headset can roughly stay on without slipping; normal, a comfortable strap length that gently secures the device; tight, shorten the strap as much as possible to fasten the device. In practice, we first ask subjects to adjust the strap to a comfortable length and gather data on the normal state. Afterward, they either loosen or tighten the strap to collect data for the other two states. The user-specific model is trained using 10 samples collected in the normal state and tested across three states. As illustrated in Fig. 15(d), HT-Auth achieves BACs of 97.33%, 96.28%, and 97.36% for normal, relaxed, and tight states, respectively. HT-Auth generally performs well under different strap



(a) (b) (c) (d)

Fig. 15. Impact of different feature sets (a), head postures (b), body postures (c), and strap tightness (d).

tightness, though in the relaxed state, the loose coupling between the VR headset and the head has a slight impact on BAC.

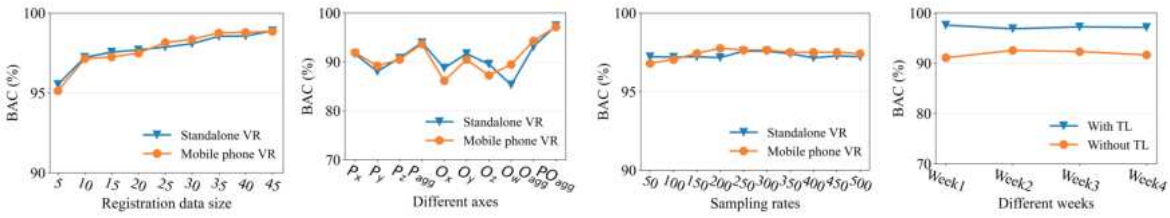
5.2.5 Human Gender. Males and Females differ in muscle morphology [48], prompting us to assess whether this affects the authentication performance. *dataset-2* is utilized for evaluation, which features both male and female subjects. Overall, our system achieves average BACs of 97.43% for males and 96.36% for females, with standard deviations of 1.71% and 1.17%, respectively. We then conduct a t-test analysis [31], which indicates that gender has no significant impact on the authentication performance, with a p-value of 0.09, exceeding the threshold of 0.05.

5.2.6 Registration Data Size. To investigate the impact of registration data size from a legitimate user, we change it from 5 to 45 in a step of 5 for fine-tuning the Siamese network. The results are shown in Fig. 16(a). As the registration data size increases from 5 to 45, the average BACs for both devices increase gradually. In particular, with only 10 registration samples, HT-Auth achieves an average BAC of 97.22% and 97.14% on standalone VR and mobile phone VR, respectively. The results demonstrate that HT-Auth can provide reliable authentication services with small registration samples.

5.2.7 Different Axes. The VR headset captures motion data across 7 axes: 3 axes for the position (P_x, P_y, P_z) and 4 axes for the orientation (O_x, O_y, O_z, O_w), each contributing unique biometric insights. We evaluate BACs for individual axes and improvements from their combination, as shown in Fig. 16(b). Note that $P_{agg} = (P_x, P_y, P_z)$, $O_{agg} = (O_x, O_y, O_z, O_w)$, and $PO_{agg} = (P_x, P_y, P_z, O_x, O_y, O_z, O_w)$. For standalone VR, the BACs for single axis range from 85.31% (O_w) to 91.69% (O_y). Aggregating multiple axes leads to additional performance gains, resulting in BACs of 93.69%, 93.12%, and 97.39% for P_{agg}, O_{agg}, and PO_{agg}, respectively. Similar trends are evident for mobile phone VR, with single-axis BACs spanning from 86.15% (O_x) to 91.90% (P_x), while aggregating axes achieve BACs of 93.65%, 94.30%, and 97.14% for P_{agg}, O_{agg}, and PO_{agg}, respectively.

5.2.8 Low Sampling Rates. To investigate the sensitivity of our system to the sensor sampling rates, we vary it from 50Hz to 500Hz in a step of 50Hz. Fig. 16(c) displays the BACs for different sampling rates across two device models. Surprisingly, we find that low sampling rates do not compromise performance. For standalone VR, the overall BAC ranges from 97.15% (400Hz) to 97.57% (250Hz), while for mobile phone VR, it fluctuates between 96.80% (50Hz) and 97.75% (200Hz). The basic reason is that head tremors typically occur within a low-frequency range (0-15Hz), making a sampling rate of 50Hz more than adequate for capturing these signals. These results indicate that our system can reliably authenticate users even at low sampling rates.

5.2.9 Consistency Over Time. To evaluate the consistency of HT-Auth over different periods, we consider three evaluation settings: i) long-term study, assessing how the gradual evolution of users' neck muscles (e.g., within a month) impacts authentication; ii) mid-term study, assessing how the natural adaptation of users' neck muscles to



(a) (b) (c) (d)

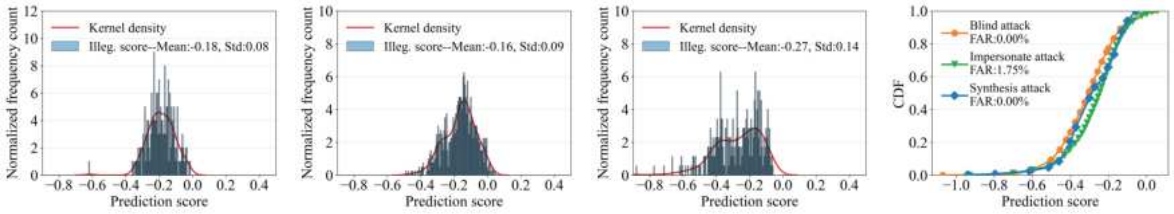
Fig. 16. BACs under different registration data sizes (a), sensor axes (b), sampling rates (c), and weeks (d).

the headset's weight and position during prolonged usage (e.g., within an hour) affects authentication; iii) short-term study, assessing system performance during consecutive authentication attempts (within a few seconds). For the long-term study, we conduct a four-week evaluation. We train the Siamese network with data from week 1 and test it with data from weeks 1, 2, 3, and 4. Fig. 16(d) illustrates the BACs throughout the four weeks. We do not observe an obvious performance decline within the four weeks. Specifically, without TL, the average BACs are 91.10%, 92.55%, 92.32%, and 91.65% for weeks 1, 2, 3, and 4, respectively. Incorporating TL boosts BACs to 97.58%, 96.86%, 97.24%, and 97.13% across the four weeks. However, when the system is used for an extended period (e.g., over a year), updating templates remains necessary, as human muscle biometrics may change with aging [7]. For the mid-term study, we consider different headset usage durations, including the initial use, 15 minutes, 30 minutes, and an hour. Subjects perform head tremors after wearing the headset for each duration. We train the Siamese network on data from the initial use and test it across different durations. The corresponding BACs are 97.25%, 96.95%, 97.18%, and 95.17%, respectively. The authentication performance remains stable during the first 30 minutes, but shows a slight decline after one hour, potentially due to muscular adaptation (i.e., subtle changes in muscle tension or activation patterns to accommodate the headset's weight) or fatigue. For the short-term study, we consider 1-5 consecutive authentication attempts and separately assess the authentication performance for users and non-users. As expected, giving users additional chances significantly improves acceptance: the FRR decreases from 3.92% on the first attempt to 0.24% after two attempts, and reaches 0% with three or more tries. For non-users, additional attempts raise the likelihood of being falsely accepted, with FARs of 1.43%, 2.64%, 4.31%, 5.64%, and 6.82% for 1-5 attempts, respectively.

5.3 Attack Resistance

5.3.1 Blind Attack. We consider blind attacks, where the attacker tries to gain access to the system through random head rotation. We choose five subjects as spoofers, asking them to wear the headset and rotate their head for two minutes. We then divide these data into 1s fragments for attacking each of the other 15 legitimate user models. We report FAR, FCS of the attack dataset, the Kernel Density (KD) of prediction scores under the Gaussian kernel [66, 69], and the Cumulative Distribution Function (CDF). Fig. 17(a), (c) displays the distribution and CDF of prediction scores for the blind attack, with the decision threshold normalized to 0. Overall, all prediction scores fall below zero, indicating that no attack samples are accepted.

5.3.2 Impersonation Attack. For the impersonation attack, we assume that the adversary is familiar with the entire authentication process through observation and aims to present similar biometric traits using their own head tremors. We utilize *dataset-2* to evaluate this attack, where each subject takes on the role of the victim, and the data from the other 19 subjects is used for attacking. 10 samples from each victim are employed for enrollment. Fig. 17(b) displays the distribution of prediction scores, where the mean and standard deviation are similar to those observed in blind attacks, with most prediction scores falling below zero. Fig. 17(d) further illustrates the



(a) Blind attack (b) Impersonation attack (c) Synthesis attack (d) CDF of prediction scores

Fig. 17. Normalized FCS/KD (a,b,c) and CDF (d) under blind, impersonation, and synthesis attacks.

CDF, revealing a low FAR of 1.75%. The results demonstrate our system's high resistance to impersonation attacks, as the unique biometrics of human neck muscles are hard for an adversary to imitate.

5.3.3 Synthesis Attack. We consider synthesis attacks, where an adversary records a victim's head tremors using cameras and attempts to reconstruct IMU signals through video processing techniques. We recruit five subjects and recorded their head tremors with two smartphones positioned approximately one meter to their front left and front right. The videos are recorded with the smartphones' rear cameras, which offer higher resolution than the front cameras, at the sampling rate of 60Hz. We employ the sparse optical flow technique [57] to quantify the displacement of head-headset contours between consecutive frames. Sparse optical flow is a computer vision method that takes a set of pixels within an image as input, typically outlining an object, and outputs a vector representing the displacement of those pixels in another image. A CNN is leveraged to perform regression, mapping head contour displacements to corresponding motion sensor readings. Specifically, the network takes two concatenated vectors (representing the head-headset contour displacements from two angles at the same timestamp) as input and outputs seven predicted values (corresponding to the 7-axis motion sensor data). It initially extracts hierarchical spatial features using four cascaded convolutional layers, flattens them into one dimension, and finally projects them into a seven-dimensional space through three dense layers to predict 7-axis motion sensor values. We train the model using data from four subjects, treating the last subject as the victim for data reconstruction. The reconstructed data is then injected into the authentication system for attack performance evaluation. Fig. 17(c) illustrates the distribution of prediction scores, where all scores fall below zero (i.e., all samples are rejected). Considering that the low sampling rate does not significantly affect authentication (see Sec. 5.2.8), we suspect the low attack performance may stem from the subtle nature of head tremors, making them challenging to capture precisely. However, if professional cameras (typically offering superior optical performance) are deployed from a wide range of angles, it is still possible for an adversary to accurately reconstruct tremor signals and bypass the authentication. Integrating the challenge-response mechanism [25, 28] can offer enhanced protection. For example, the system can ask the user to perform tremors following a randomly generated rhythm, guided visually or audibly. The rhythm is first verified to confirm liveness, and then the biometric information is utilized for identity determination.

5.4 Evaluation of Computational Delay

In practical user authentication scenarios, short inference time is crucial for enabling real-time authentication and better user experience [60, 67, 68]. To demonstrate that HT-Auth is suitable for user authentication in practical VR environments, we assess the average computational time across its various modules. In particular, we conduct experiments using a Tesla P4 GPU and Intel i7-9700 CPU with a batch size of 8 for 1000 samples to measure the average computational time for noise reduction, event detection, level-I and II feature extraction, Siamese network-based feature reconstruction, and user authentication. We summarize the average computational

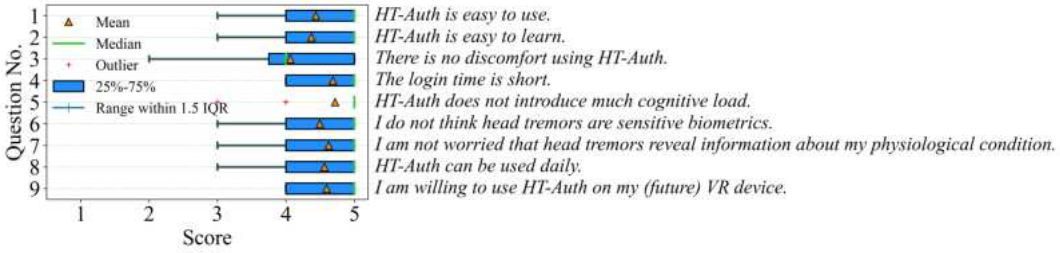


Fig. 18. Results of the user study questionnaire.

times for various modules within HT-Auth. The event detection has the longest processing time, averaging 35.28ms. For noise reduction, level-I and II feature extraction, feature reconstruction, and user authentication, the average computational times are 5.24ms, 20.35ms, 11.36ms, and 2.47ms, respectively, which indicates that HT-Auth processes a single tremor input and completes user authentication in approximately 74.70ms. The short computational time suggests that HT-Auth can be deployed in practical VR environments, enabling real-time user authentication.

5.5 User Study

We conduct a user study to assess the subjects' subjective perception of HT-Auth. After data collection, all 30 subjects are asked to provide their perspective of HT-Auth by responding to 9 questions on a 5-point Likert scale, where 1 means "strongly disagree" and 5 means "strongly agree". Fig. 18 presents the questions and the corresponding results. Most subjects consider that HT-Auth is a viable daily authentication solution for VR devices (Q8) and express a strong willingness to use it (Q9), with mean scores (μ) of 4.56 and 4.59, and median scores (m) of 5 for both questions. In terms of system usability, the responses to all five related questions are positive: Q1 ($\mu = 4.44, m = 5$), Q2 ($\mu = 4.38, m = 5$), Q3 ($\mu = 4.06, m = 4$), Q4 ($\mu = 4.69, m = 5$), and Q5 ($\mu = 4.72, m = 5$). Regarding privacy concerns, most subjects do not view head tremors as sensitive biometrics and are not concerned that they might disclose information about their physiological conditions, with Q6 ($\mu = 4.50, m = 5$) and Q7 ($\mu = 4.61, m = 5$). Overall, HT-Auth is well recognized by subjects (Q8, Q9), due to its ease of use (Q1), ease of learning (Q2), no discomfort (Q3), short login time (Q4), low cognitive load (Q5), and less sensitive (Q6, Q7).

Apart from the above questions, we also verbally ask subjects if they have any concerns about the authentication system capturing, storing, and processing their head tremor patterns. 2 subjects express behavioral privacy concerns, e.g., *"Can my tremor behavior be observed externally? Could an attacker potentially mimic it?"* Although head tremors are subtle, they can still be observed externally. In Sec. 5.3, we demonstrate that it is difficult for attackers to bypass authentication by mimicking the victim's behavior or synthesizing data from video recordings. 1 subject expresses data privacy concerns, e.g., *"How is my data stored? Is it encrypted?"* Specifically, the user's biometric template will be stored to match the registered data. In HT-Auth, the template is the embedding vector R, rather than the raw tremor data. To enhance privacy protection, R can be stored encrypted and only decrypted briefly for sample matching during authentication.

6 RELATED WORK

In this section, we first discuss existing authentication methods designed for VR environments. Then we present the related works that utilize muscle biometrics for user authentication.

6.1 User Authentication on AR/VR

While passwords and PINs serve as the most popular authentication methods on AR/VR devices [1], they are criticized for these usability issues: i) using a virtual laser extended from the controller to interact with the virtual keyboard is slow, difficult, and prone to input errors [58]; ii) the input actions in the physical world can be observed by surrounding people or cameras, rendering it possible for an attacker to infer the entered credentials [19]. To address these issues, biometrics-based authentication methods have been proposed as practical alternatives. These methods primarily fall into two categories: *behavior biometrics-based* and *physiological biometrics-based*.

Behavior biometrics-based methods authenticate users based on unique user behavior patterns. For example, head movements have been proven to effectively identify users when they listen to music beats [38], moving the VR pointer to follow/throw a ball [45, 47], or walking [53]. Moreover, the VR headset can work with handheld controllers, capturing not only head movements, but also hand movements, body movements and even eye gaze when users perform required tasks in the virtual space [39, 49, 50, 76]. Face-Mic [54] and SAFARI [75] leverage subtle facial movement during speaking to identify users, where the facial movement patterns are identified. Mathis et al. [43, 44] ask users to enter numeric passwords by manipulating a 3D cube in the virtual space, where both the passwords and handheld controller motions are verified. However, most of these methods are inefficient, requiring a relatively long time to complete the task.

Physiological biometrics-based methods leverage invariable body traits to identify users. These body traits could be electrooculogram (EOG) [41] and electroencephalogram (EEG) [40] responses to visual stimuli, or electromyogram (EMG) [10] responses to electrical impulses. However, dedicated sensors are required for these methods. SoundLock [78] recognizes users by capturing the auditory-pupillary response as biometrics. But it requires a relatively long time for user enrollment (800 to 820 seconds), and many VR headsets lack a built-in eye tracker, making it inefficient to use. To achieve low-effort authentication, some studies utilize head-conducted vibrations [36] and acoustic signals [52, 65] for biometric measurement. However, these methods rely on extra challenge signals, such as vibrations or sound chirps, which can be intrusive and often impractical without hardware modifications, given the default echo cancellation mechanism of VR devices. A comparison of HT-Auth with representative VR-based authentication methods is presented in Table 2. Overall, compared with these studies, HT-Auth requires no dedicated hardware, delivers commendable performance with a small registration size, and has been evaluated across extensive settings.

6.2 Muscle Biometrics for User Authentication

The muscle biometrics are relatively accessible and hard-to-spoof traits for user authentication [63]. Most studies characterize these biometrics by measuring the EMG signals generated by user activities, such as keystroke dynamics [63] or hand gestures [74]. The EMG can be integrated with other signals, e.g., electrocardiograph (ECG), to generate more reliable authentication results [5]. Chen et al. [10] leverage electrical impulses to actively stimulate the user's forearm muscles, measuring the user's involuntary finger movements for authentication. Ataş et al. [4] utilize leap motion devices to capture hand tremor as biometrics. However, these methods all require dedicated hardware and cannot be deployed on commodity VR devices. MAUTH [29] leverages motion sensors on smartphones to extract hand tremors for continuous authentication. However, due to the high variability in the user's handheld postures, its performance degrades significantly in a short time. e.g., decrease by 15% after four days. Differently, our work authenticates users by measuring neck muscular biometrics through unique head tremors, which is compatible with mainstream VR headsets for achieving secure authentication.

7 DISCUSSION

Information acquisition and spoof execution. In practice, adversaries may obtain the victim's authentication information and execute spoofing at different stages of the authentication flow. Information can be obtained

Table 2. A comparison with representative VR-based authentication methods.

Work	Modality	Dedicated sensor	Registration size[1]	Performance	Evaluation settings[2]				
					Different devices	Head/body postures	Template transferability	Time consistency	User study
GaitLock [53]	Gait	×	60(~30s)	~98% Acc.	×	✓	×	✓	×
SAFARI [75]	Facial deformation	×	200(~400s)	95.71% BAC	✓	✓	×	×	×
OcuLock [41]	EOG	✓	N/A	3.55% EER	×	×	×	✓	✓
Lin et al. [40]	EEG	✓	270(54s)	2.50% EER	×	×	×	✓	×
Chen et al. [10]	EMG	✓	920(1140s)	99.78% Acc.	×	×	×	✓	×
SoundLock [78]	Pupil deformation	×	N/A(~810s)	1.50% EER	×	✓	×	✓	✓
HT-Auth (ours)	Head tremor	×	10(~10s)	97.22% BAC	✓	✓	✓	✓	✓

[1] Registration size is the number of samples per subject needed for model training to achieve the reported performance. To enable fair comparison across studies, we include the data collection time of these samples.
[2] For each evaluation setting, we considered multiple aspects. If an existing study covers at least one aspect or has considered a similar situation, we regard it as having studied that setting and mark it with a ✓.

Table 3. The attack methods evaluated (✓) or discussed (○) in this study.

Info acquisition ↓ / Spoof execution →	Human head	Dummy robot head	Data injection
Zero information	Blind attack (✓)	-	-
Human observation	Impersonation attack (✓)	-	-
Camera recording	-	-	Synthesis attack (✓)
Malicious app	-	Puppetry attack (○)	Injection attack (○)

in the physical world or by deploying malicious software. Specifically, the simplest method involves watching the victim's head tremors, either through direct observation or camera recording. A skilled adversary can trick the victim into installing a malicious app on the target VR device. Once installed, it silently gathers motion sensor data during the victim's tremor activity, exploiting the fact that such sensors typically operate with zero permissions [27]. To bypass authentication, spoofing can also be executed at either the physical or software levels. In the physical world, for instance, an adversary can perform head tremors by mimicking the victim's behavior with a human or dummy robot head, without compromising the authentication system's integrity. The dummy robot head relies on precise servo motor control to replicate the victim's behavior, typically involving high costs. A skilled adversary can directly inject the victim's data into the authentication system. The injection can be done by either tampering with the motion sensors or manipulating the sensor APIs. For example, the default motion sensor can be replaced with a manipulated one to produce the injected data. By embedding unique hardware identifiers for sensors and verifying them during data transmission, such tampering can be mitigated. On the other hand, applications access motion sensor data via Android system APIs on VR devices. To inject data, the adversary can compromise the transmission path, replacing the default API data with malicious input. This typically requires rooting the device or manipulating system drivers, both of which present significant challenges. We summarize the attacks corresponding to different information acquisition and spoof execution strategies in Table 3.

Advanced attack resistance. An adversary may launch advanced attacks by obtaining the victim's authentication data via malicious software. Spoofing can then be executed using either a servo motor-controlled dummy robot head (puppetry attack) or malicious software-controlled APIs (injection attack). These attacks, though complex for adversaries, pose a serious challenge to our authentication system as they can precisely replicate the victim's biometric information. An intuitive defense is to restrict the motion sensor usage. For example, they can be treated as seriously as cameras, requiring explicit user permission before any access. Since these attacks typically follow a pre-defined execution flow, integrating a challenge-response mechanism [25, 28] can offer extra protection, as detailed in Sec. 5.3.3. Moreover, liveness detection can serve as a simple solution to counteract these attacks. For

instance, the system can prompt users to blink (many consumer VR headsets are equipped with eye trackers), nod, or shake their head before performing head tremors. Users are granted access only after successfully passing both liveness detection and the proposed authentication system.

Consecutive authentication attempts. As illustrated in Sec. 5.2.9, users achieve near-perfect authentication within 2–3 attempts, whereas non-users show a steadily increasing risk of bypassing security as the number of attempts increases, with the FAR rising from 1.43% to 6.82%. This reveals a critical trade-off between usability and security: allowing unlimited retries benefits users little after a few attempts but substantially weakens system security. To address this trade-off, a practical strategy is to cap consecutive authentication attempts ((e.g., 2–3 tries)) within a given period, after which users are seamlessly redirected to fallback methods such as password authentication.

Active and passive tremors. This study focuses on active tremor, which arises when users willingly tense their muscles. Due to the effort required to sustain it, this tremor typically lasts for a short duration (around 1s) for muscular fatigue prevention. In fact, even in a relaxed state, the user's muscles are not completely still, resulting in passive tremors. Passive tremors are less stable and more susceptible to user habits. For instance, a prior study using passive hand tremors for continuous authentication exhibits a 13% drop in EER over four days [29]. In contrast, active tremors can maintain consistency over much longer periods—for example, sustaining commendable authentication performance over four weeks, as demonstrated in Sec. 5.2.9. Overall, passive tremors do not affect our authentication approach. Since these tremors resemble "stillness" more than "movement", they will not be considered valid events by our ASV-based event detection algorithm.

8 LIMITATIONS AND FUTURE WORK

Considering that the proposed authentication system still exhibits a non-negligible FAR (1.75% under impersonation attacks), it may not be suitable for certain security-critical scenarios, such as financial or healthcare applications. Instead, HT-Auth is more suitable for non-critical scenarios, such as home device activation, sport or game account login, and personalized environment or interface customization. HT-Auth requires active head tremors, and repeated attempts in quick succession may lead to user fatigue. The system's performance remains affected by head poses and human motions. The collected datasets involved a limited number of subjects, with demographic factors such as gender, region, and age not fully balanced. The standalone headset studied in this work is the Meta Quest 2, chosen for its low prices and high shipments. In future work, we plan to conduct evaluations on a larger subject dataset, covering a wide range of races, ages, weights, and body fat ratios. Future work will also include evaluating additional standalone VR devices, such as the HTC VIVE Pro and HP Reverb, as well as investigating the feasibility of applying our system to AR platforms.

9 CONCLUSION

This paper introduces HT-Auth, a secure authentication method for VR devices based on subtle head tremors. HT-Auth utilizes an inertial sensor built-in VR headset to enable its functionality. In particular, we design a series of techniques for pre-processing the raw tremor signals, including PSD analysis, GA-based MODWT denoising, and a three-stage event localization algorithm to filter out body motion interference and provide signal localization. We also design a novel biometric representation that characterizes the unique contraction and endurance characteristics of human neck muscles. Furthermore, we introduce a Siamese network-based transfer learning algorithm to authenticate newly registered users efficiently. Extensive experiments show that HT-Auth achieves a 97.22% BAC to distinguish users with as few as 10 registration samples and can defend against blind and impersonation attacks.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under grants No. 62172303, 62472323, the Key R&D Program of Hubei Province under grant No. 2024BAB018, Wuhan Scientific and Technical Achievements Project under Grant No.2024030803010172, and the Key R&D Program of Shandong Province under grant No. 2022CXPT055. The corresponding author is Jing Chen.

References

[1] 2024. *Get started with Meta Quest 2.* https://www.meta.com/au/quest/products/quest-2/.
[2] 2024. *Report: volume of the VR headsets market.* https://www.statista.com/forecasts/1331896/vr-headset-sales-volume-worldwide.
[3] Zaid Abdi Alkareem Alyasseri, Ahamad Tajudin Khader, Mohammed Azmi Al-Betar, Ammar Kamal Abasi, and Sharif Naser Makhadmeh. 2019. EEG signals denoising using optimal wavelet transform hybridized with efficient metaheuristic methods. *IEEE Access* (2019).
[4] Musa Ataş. 2017. Hand tremor based biometric recognition using leap motion device. *IEEE Access* (2017).
[5] Noureddine Belgacem, Régis Fournier, Amine Nait-Ali, and Fethi Bereksi-Reguig. 2015. A novel biometric authentication approach using ECG and EMG signals. *Journal of medical engineering & technology* (2015).
[6] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. 2010. The balanced accuracy and its posterior distribution. In *IEEE conference on pattern recognition (ICPR)*.
[7] Gregory D Cartee, Russell T Hepple, Marcas M Bamman, and Juleen R Zierath. 2016. Exercise promotes healthy aging of skeletal muscle. *Cell metabolism* (2016).
[8] Wenqiang Chen, Lin Chen, Yandao Huang, Xinyu Zhang, Lu Wang, Rukhsana Ruby, and Kaishun Wu. 2019. Taprint: Secure text input for commodity smart wristbands. In *ACM International Conference on Mobile Computing and Networking (MobiCom)*.
[9] Yanjiao Chen, Meng Xue, Jian Zhang, Qianyun Guan, Zhiyuan Wang, Qian Zhang, and Wei Wang. 2021. Chestlive: Fortifying voice-based authentication with chest motion biometric on smart devices. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2021).
[10] Yuxin Chen, Zhuolin Yang, Ruben Abbou, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2021. User authentication via electrical muscle stimulation. In *CHI Conference on Human Factors in Computing Systems (CHI)*.
[11] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[12] Constantinos N Christakos, Nikos A Papadimitriou, and Sophia Erimaki. 2006. Parallel neuronal mechanisms underlying physiological force tremor in steady muscle contractions of humans. *Journal of neurophysiology* (2006).
[13] Charles R Cornish, Christopher S Bretherton, and Donald B Percival. 2006. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology* (2006).
[14] DL Costill, EF Coyle, WF Fink, GR Lesmes, and FA Witzmann. 1979. Adaptations in skeletal muscle following strength training. *Journal of Applied Physiology* (1979).
[15] Dian Ding, Lanqing Yang, Yi-Chao Chen, and Guangtao Xue. 2021. Leakage or identification: Behavior-irrelevant user identification leveraging leakage current on laptops. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2021).
[16] Ceenu George, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. Network and Distributed System Security Symposium (NDSS).
[17] Anne M Gilroy, Brian R MacPherson, Lawrence M Ross, Jonas Broman, and Anna Josephson. 2008. *Atlas of anatomy*. Thieme Stuttgart.
[18] Ph D Gollnick. 1982. Relationship of strength and endurance with skeletal muscle structure and metabolic potential. *International Journal of Sports Medicine* (1982).
[19] Sindhu Reddy Kalathur Gopal, Diksha Shukla, James David Wheelock, and Nitesh Saxena. 2023. Hidden reality: Caution, your hand gesture inputs in the immersive virtual world are visible to all!. In *USENIX Security Symposium (USENIX)*.
[20] Yangyang Gu, Jing Chen, Congrui Chen, Kun He, Jia Ju, Yebo Feng, Ruiying Du, and Cong Wu. 2025. CSIPose: Unveiling Human Poses Using Commodity WiFi Devices Through the Wall. *IEEE Transactions on Mobile Computing (TMC)* (2025).
[21] Yangyang Gu, Jing Chen, Kun He, Cong Wu, Ziming Zhao, and Ruiying Du. 2023. Wifileaks: Exposing stationary human presence through a wall with commodity mobile devices. *IEEE Transactions on Mobile Computing (TMC)* (2023).
[22] Yangyang Gu, Jing Chen, Cong Wu, Kun He, Ziming Zhao, and Ruiying Du. 2024. Loccams: An efficient and robust approach for detecting and localizing hidden wireless cameras via commodity devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2024).

[23] Feiyu Han, Panlong Yang, Shaojie Yan, Haohua Du, and Yuanhao Feng. 2023. BreathSign: Transparent and continuous in-ear authentication using bone-conducted breathing biometrics. In *IEEE International Conference on Computer Communications (INFOCOM)*.
[24] Zhixiang He, Jing Chen, Kun He, Yangyang Gu, Qiyi Deng, Zijian Zhang, Ruiying Du, Qingchuan Zhao, and Cong Wu. 2025. HeadSonic: Usable Bone Conduction Earphone Authentication via Head-conducted Sounds. *IEEE Transactions on Mobile Computing (TMC)* (2025).
[25] Zhixiang He, Jing Chen, Kun He, Cong Wu, Xiangyu Qu, Yangyang Gu, Xiping Sun, and Ruiying Du. 2025. EyeAuth: smartphone user authentication via reflexive eye movements. *Frontiers of Computer Science (FCS)* (2025).
[26] Zhixiang He, Jing Chen, Cong Wu, Kun He, Ruiying Du, Ju Jia, Yangyang Gu, and Xiping Sun. 2024. HCR-Auth: Reliable Bone Conduction Earphone Authentication with Head Contact Response. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2024).
[27] Pengfei Hu, Hui Zhuang, Panneer Selvam Santhalingam, Riccardo Spolaor, Parth Pathak, Guoming Zhang, and Xiuzhen Cheng. 2022. Accear: Accelerometer acoustic eavesdropping with unconstrained vocabulary. In *IEEE Symposium on Security and Privacy (SP)*.
[28] Long Huang and Chen Wang. 2022. Pcr-auth: Solving authentication puzzle challenge with encoded palm contact response. In *IEEE Symposium on Security and Privacy (SP)*.
[29] Yi Jiang, Hongzi Zhu, Shan Chang, and Bo Li. 2023. MAUTH: Continuous user authentication based on subtle intrinsic muscular tremors. *IEEE Transactions on Mobile Computing (TMC)* (2023).
[30] Rinat Khusainov, Djamel Azzi, Ifeyinwa E Achumba, and Sebastian D Bersch. 2013. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors* (2013).
[31] Tae Kyun Kim. 2015. T test as a parametric statistic. In *Korean journal of anesthesiology*.
[32] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014).
[33] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. 2015. Siamese neural networks for one-shot image recognition. In *International Conference on Machine Learning (ICML)*.
[34] David M Kreindler and Charles J Lumsden. 2016. The effects of the irregular sample and missing data in time series analysis. In *Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data*.
[35] Christopher M Laine, Akira Nagamori, and Francisco J Valero-Cuevas. 2016. The dynamics of voluntary force production in afferented muscle influence involuntary tremor. *Frontiers in computational neuroscience* (2016).
[36] Feng Li, Jiayi Zhao, Huan Yang, Dongxiao Yu, Yuanfeng Zhou, and Yiran Shen. 2024. Vibhead: An authentication scheme for smart headsets through vibration. *ACM Transactions on Sensor Networks* (2024).
[37] Jingjie Li, Kassem Fawaz, and Younghyun Kim. 2019. Velody: Nonlinear vibration challenge-response for resilient user authentication. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*.
[38] 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 *IEEE International Conference on Pervasive Computing and Communications (PerCom)*.
[39] Jonathan Liebers, Patrick Laskowski, Florian Rademaker, Leon Sabel, Jordan Hoppen, Uwe Gruenefeld, and Stefan Schneegass. 2024. Kinetic signatures: A systematic investigation of movement-based user identification in virtual reality. In *CHI Conference on Human Factors in Computing Systems (CHI)*.
[40] Feng Lin, Kun Woo Cho, Chen Song, Wenyao Xu, and Zhanpeng Jin. 2018. Brain password: A secure and truly cancelable brain biometrics for smart headwear. In *International Conference on Mobile Systems, Applications, and Services (Mobisys)*.
[41] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenyao Xu, and Zhisheng Yan. 2020. OcuLock: Exploring human visual system for authentication in virtual reality head-mounted display. In *Network and Distributed System Security Symposium (NDSS)*.
[42] Stephane G Mallat. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)* (1989).
[43] Florian Mathis, Hassan Ismail Fawaz, and Mohamed Khamis. 2020. Knowledge-driven biometric authentication in virtual reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*.
[44] Florian Mathis, John Williamson, Kami Vaniea, and Mohamed Khamis. 2020. Rubikauth: Fast and secure authentication in virtual reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*.
[45] Robert Miller, Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Realtime behavior-based continual authentication of users in virtual reality environments. In *International Conference on Artificial Intelligence and Virtual Reality (AIVR)*.
[46] Meinard Müller. 2007. Dynamic time warping. *Information retrieval for music and motion* (2007).
[47] Tahrima Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. Unsure how to authenticate on your vr headset? come on, use your head!. In *ACM International Workshop on Security and Privacy Analytics*.
[48] Margareta Nordin. 2020. *Basic biomechanics of the musculoskeletal system*. Lippincott Williams & Wilkins.
[49] Ilesanmi Olade, Charles Fleming, and Hai-Ning Liang. 2020. Biomove: Biometric user identification from human kinesiological movements for virtual reality systems. *Sensors* (2020).

[50] Ken Pfeuffer, Matthias J Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural biometrics in VR: Identifying people from body motion and relations in virtual reality. In *CHI Conference on Human Factors in Computing Systems (CHI)*.
[51] Olivier Rukundo and Hanqiang Cao. 2012. Nearest neighbor value interpolation. *arXiv preprint arXiv:1211.1768* (2012).
[52] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *CHI Conference on Human Factors in Computing Systems (CHI)*.
[53] 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 (TDSC)* (2018).
[54] Cong Shi, Xiangyu Xu, Tianfang Zhang, Payton Walker, Yi Wu, Jian Liu, Nitesh Saxena, Yingying Chen, and Jiadi Yu. 2021. Face-Mic: inferring live speech and speaker identity via subtle facial dynamics captured by AR/VR motion sensors. In *ACM Conference on Mobile Computing and Networking (MobiCom)*.
[55] Dai Shi, Dan Tao, Jiangtao Wang, Muyan Yao, Zhibo Wang, Houjin Chen, and Sumi Helal. 2021. Fine-grained and context-aware behavioral biometrics for pattern lock on smartphones. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2021).
[56] Paul F Sowman and Kemal S Türker. 2005. Methods of time and frequency domain examination of physiological tremor in the human jaw. *Human movement science* (2005).
[57] David Stavens. 2007. The OpenCV library: computing optical flow.
[58] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. 2022. Sok: Authentication in augmented and virtual reality. In *IEEE Symposium on Security and Privacy (SP)*.
[59] Shridatt Sugrim, Can Liu, Meghan McLean, and Janne Lindqvist. 2019. Robust performance metrics for authentication systems. In *In Network and Distributed Systems Security Symposium (NDSS)*.
[60] Xiping Sun, Jing Chen, Kun He, Zhixiang He, Ruiying Du, Yebo Feng, Qingchuan Zhao, and Cong Wu. 2025. SCR-Auth: Secure Call Receiver Authentication on Smartphones Using Outer Ear Echoes. *IEEE Transactions on Information Forensics and Security (TDSC)* (2025).
[61] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research (JMLR)*. (2008).
[62] J Vartiainen, JJ Lehtomaki, and H Saarnisaari. 2005. Double-threshold based narrowband signal extraction. In *IEEE vehicular technology conference*.
[63] Shreyas Venugopalan, Felix Juefei-Xu, Benjamin Cowley, and Marios Savvides. 2015. Electromyograph and keystroke dynamics for spoof-resistant biometric authentication. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
[64] Dhruv Verma, Sejal Bhalla, Dhruv Sahnan, Jainendra Shukla, and Aman Parnami. 2021. Expressear: Sensing fine-grained facial expressions with earables. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2021).
[65] Ruxin Wang, Long Huang, and Chen Wang. 2023. Low-effort VR Headset User Authentication Using Head-reverberated Sounds with Replay Resistance. In *IEEE Symposium on Security and Privacy (SP)*.
[66] Cong Wu, Jing Chen, Kun He, Ziming Zhao, Ruiying Du, and Chen Zhang. 2022. EchoHand: High Accuracy and Presentation Attack Resistant Hand Authentication on Commodity Mobile Devices. In *ACM conference on computer and communications security (CCS)*.
[67] Cong Wu, Kun He, Jing Chen, and Ruiying Du. 2019. ICAuth: Implicit and continuous authentication when the screen is awake. In *IEEE International Conference on Communications (ICC)*.
[68] Cong Wu, Kun He, Jing Chen, Ruiying Du, and Yang Xiang. 2020. CaiAuth: Context-aware implicit authentication when the screen is awake. *IEEE Internet of Things Journal (IoT-J)* (2020).
[69] Cong Wu, Kun He, Jing Chen, Ruiying Du, Ran Yan, and Ziming Zhao. 2025. High Accuracy and Presentation Attack Resistant Hand Authentication via Acoustic Sensing for Commodity Mobile Devices. *IEEE Transactions on Dependable and Secure Computing (TDSC)* (2025).
[70] Cong Wu, Kun He, Jing Chen, Ziming Zhao, and Ruiying Du. 2020. Liveness is not enough: Enhancing fingerprint authentication with behavioral biometrics to defeat puppet attacks. In *USENIX Security Symposium (USENIX)*.
[71] Cong Wu, Kun He, Jing Chen, Ziming Zhao, and Ruiying Du. 2021. Toward robust detection of puppet attacks via characterizing fingertip-touch behaviors. *IEEE Transactions on Dependable and Secure Computing (TDSC)* (2021).
[72] Xiangyu Xu, Jiadi Yu, Yingying Chen, Qin Hua, Yanmin Zhu, Yi-Chao Chen, and Minglu Li. 2020. TouchPass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations. In *ACM International Conference on Mobile Computing and Networking (MobiCom)*.
[73] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Annual international conference on mobile systems, applications, and services*.
[74] Lin Yang, Wei Wang, and Qian Zhang. 2016. Secret from muscle: Enabling secure pairing with electromyography. In *ACM Conference on Embedded Network Sensor Systems CD-ROM*.

[75] Tianfang Zhang, Qiufan Ji, Zhengkun Ye, Md Mojibur Rahman, Redoy Akanda, Ahmed Tanvir Mahdad, Cong Shi, Yan Wang, Nitesh Saxena, and Yingying Chen. 2024. SAFARI: Speech-Associated Facial Authentication for AR/VR Settings via Robust VIbration Signatures. In *ACM SIGSAC conference on computer and communications security (CCS)*.
[76] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous authentication using eye movement response of implicit visual stimuli. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2018).
[77] Huadi Zhu, Wenqiang Jin, Mingyan Xiao, Srinivasan Murali, and Ming Li. 2020. Blinkey: A two-factor user authentication method for virtual reality devices. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMUWT)* (2020).
[78] Huadi Zhu, Mingyan Xiao, Demoria Sherman, and Ming Li. 2023. SoundLock: A Novel User Authentication Scheme for VR Devices Using Auditory-Pupillary Response.. In *Network and Distributed System Security Symposium (NDSS)*.