Exploring a Brain-Based Cancelable Biometrics for Smart Headwear: Concept, Implementation, and Evaluation

Feng Lin, *Member, IEEE*, Kun Woo Cho, *Student Member, IEEE*, Chen Song, *Student Member, IEEE*, Zhanpeng Jin, *Senior Member, IEEE*, and Wenyao Xu, *Senior Member, IEEE*

Abstract—Biometric authentication offers advantages over current security practices. Unlike keys and tokens, biometrics are never lost or stolen. Unlike passwords, biometrics cannot be forgotten. However, existing biometric systems are with controversy: once divulged, they are compromised forever. To this end, this paper explores a truly cancelable brain-based biometric system for the first time. Specifically, we present a new psychophysiological protocol via non-volitional brain response for trustworthy user authentication, with an application example of smart headwear. More specifically, we address the following research challenges in a theoretical and experimental combined manner: (1) how to generate reliable brain responses with sophisticated visual stimuli; (2) how to acquire effective brain response and analyze unique features in them for authentication; and (3) how to reset and change brain biometrics when the current biometric credential is divulged. To evaluate the performance of the proposed system, we conducted a pilot study and achieved an f-score accuracy of 95.46 percent and equal error rate (EER) of 2.503 percent, thereby demonstrating the potential feasibility of neurofeedback based biometrics for smart headwear applications. Further, the cancelability study proves the effectiveness of the reset brain password. To the best of our knowledge, it is the first in-depth research study on truly cancelable brain biometrics.

Index Terms—Head-mounted display, event-related potential (ERP), cancelable brain biometrics, mobile authentication

1 Introduction

In recent years, biometric authentication is taking over traditional passwords or PIN based authentication in mobile and wearable applications because of identification accuracy, convenience and seamless integration with personal devices. However, existing biometrics, such as fingerprint and face, are prone to being hacked in everyday life or social media. For example, the Chaos Computer Club announced that one of its members had been able to replicate the fingerprint of German Minister of Defense Ursula von der Leyen, using only photographs taken of her finger [1]. Biometrics are unique to individual. Different from traditional passwords, once such biometric credentials are damaged or counterfeited, the user cannot cancel the pre-stored credentials or reset them with a different biometric input.

How to design a *truly cancelable* biometric system is a long and historical topic in the biometric research community. Cancelable biometrics are challenging because stability and cancelability in biometrics are at odds with each other. Stability requires that biometric traits are immutable and hard to change; cancelability require that biometric traits are erasable and easy to change. According to our literature review, existing works on cancelable biometrics mainly focus on "soft cancellation", which means the biometric system only uses and saves transformed biometric credentials, such as images with random projection, in the database. In other words, once biometric information is divulged, users can generate a new biometric template with a different transformation approach. For example, Paul et al. [2] introduced a cancelable template generation algorithm, when previously transformed template is stolen, that produces new transformed biometric templates. The algorithm can make new templates unlinkable to the previous compromised template. Nevertheless, this soft-cancellation approach only works in case of database breaches. If the original biometrics are compromised (e.g., compromising raw fingerprint patterns in a photograph), it still results in permanent biometric compromise. Therefore, to address the fundamental limitation of biometrics, it is necessary to explore a new biometric approach for "hard cancellation".

In recent years, physiological activities from human organs (e.g., brains [3]) receive increasing attention in biometric communities. The advantage of brain electric activity based biometrics is that they are biologically unique and less prone to forgery because of the dynamics of brain responses. For example, event-related potential (ERP) brainwave is one type of brain electrical signals and can be changed once different visual stimuli are presented [4]. This special feature of brain response offers the potential to

- *F. Lin is with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027, China. E-mail: flin@zju.edu.cn.*
- *K.W. Cho, C. Song, Z. Jin, and W. Xu are with the Department of Computer Science & Engineering, University at Buffalo (SUNY), Buffalo, NY 14260. E-mail: {kunwooch, csong5, zjin, wenyaoxu}@buffalo.edu.*

Manuscript received 21 May 2018; revised 1 June 2019; accepted 9 Aug. 2019. Date of publication 20 Aug. 2019; date of current version 3 Nov. 2020. (Corresponding author: Wenyao Xu.)
Digital Object Identifier no. 10.1109/TMC.2019.2936559

design a truly cancelable biometrics, referring to "hard-cancellation". For example, if an ERP brainwave is produced in response to a series of images, that ERP brainwave can be canceled, and a new ERP brainwave can be generated in response to another series of image stimuli.

Here, we argue that the most secure cryptographic credential can be obtained by ERP brainwave signals. By definition, ERP is one of the brain biometric measures that are related to individual-specific characteristics. Besides its unique property of hard-cancellation, ERP also possesses another superior attribute compared with traditional biometrics. While conventional anatomical and behavioral biometrics, such as a fingerprint, voice, stroke, and gait, are not confidential to an individual or can easily be altered for imitation [5], [6], ERP biometrics are highly secure; one cannot reproduce or copy other person's mental pass-phrase. Moreover, it is non-revealable and naturally less prone to spoofing and counterfeiting [7]. In summary, the ERP-based brainwave biometrics stand out with the following advantages:

- *Secure*: Traditional brainwaves biometrics which requires users to create thought patterns to generate the corresponding brainwave credentials [8]. In this case, brainwave credentials are consciously controlled by users, which can be revealed either purposely or unintentionally [9]. Instead, ERP is the non-volitional and involuntary brain response. This mechanism conceals conspicuous interactions and provides better security, i.e., a user himself even has no control of ERP generations.
- *Cancelable*: Part of what makes each brain unique is their knowledge and memories. The brain network that manages forming and accessing memories is large, spans across many anatomical areas [10]. This provides a potential large capacity of various brain ERP responses. Therefore, if ERP template database is breached, new user's ERP credentials are possible to be generated by different stimuli sets. Notes that ERP biometrics also requires no memorization burden on users as other traditional biometrics (e.g., PIN, graphical pattern).

Based on these arguments, we propose a new psychophysiological approach for secure and trustworthy user authentication in a head-mounted display (HMD). An HMD is a computerized, information viewing device that is worn on the head. It consists a small display optic in front of eyes, which covers the entire field of view of the user and produces an imaginary screen that appears to be positioned away from eyes. Since both ERP acquisition sensor and HMD are mounted on the head (see Section 3 later), it is natural to employ ERP biometrics for the authentication of smart headwear.

In this work, we study ERP, a non-volitional and involuntary brainwave response to a specific sensory, cognitive, or visual stimulation, for HMD authentication. To generate distinct ERP patterns for biometric applications, we utilize a visual stimuli design consisting of the imagery patterns of animal, human face, and text as examples. The brain activity data are acquired from a lightweight wearable brain-computer interface with three channels (i.e., P1, Pz, and P4), and the features are extracted from multiple models, such as an autoregressive (AR) model, power spectral density (PSD), and eigenvector. Then, the feature vectors are classified via support vector machine. Our main challenge is when the ERP credential is divulged, what is the effective strategy to reset and generate new and secure ERPs. In this study, we present a novel stimuli update strategy that updates the in-use stimuli to evoke new stable ERPs. In analogy to the case where the user is not allowed to use a password that is too close to a previous selection, we characterize the sequence of visual stimuli in a joint spatio-temporal domain and choose the ERP with the maximum proposed spatio-temporal warping distance as the new credential. As a result, the original and newly generated "brain passwords" are disparate that the original ERP cannot be cross-matched to access the system configured with new ERP credentials. Also, the system maintains stability in this way as the new ERP retains immutability until it is divulged again. To validate the proposed approach, we further conduct a pilot study to evaluate the system security via f-score accuracy (f-1), half total error rate (HTER), receiver operating characteristic curve (ROC), equal error rate (EER), and the time efficiency. With 179 adult participants, our system achieves the f-score accuracy of 95.46 percent, HTER of 2.261 percent, and EER of 2.503 percent. The cancelability evaluation proves that our stimuli update strategy is effective in revoking old ERP and reissuing new ERP derived from the same physical traits without degrading the authentication performance. Also, the unlinkability between old and new ERPs is discussed in this study.

To the best of our knowledge, this is the first in-depth study to explore secure and *truly cancelable* biometrics for user authentication. In sum, there are three contributions in our work:

- We propose a secure and truly cancelable ERP-based authentication protocol with its application for smart headwears. We study a sophisticated brain response model and develop an end-to-end brain biometric system integrated with a head-worn device.
- We study a joint spatio-temporal domain analysis-based stimuli update strategy to achieve the cancelability of our proposed biometric protocol. We empirically investigate the biometric capacity of brain response.
- We validate the feasibility and effectiveness of our proposed system with multi-session pilot studies, including the performance study, cancelability study and longitudinal study in different user scenarios.

2 Background and Theory

2.1 HMD Authentication

2.1.1 Significance of HMD

In recent years, HMDs have been widely developed and improved for a variety of purposes. Main applications include virtual reality (VR) for simulation of user's presence in artificial environments (Samsung VR [11]) and realistic experience of 3D games (PlayStation VR [12]). Also, some HMDs provide an augmented reality (AR) to integrate digital information with user's real-world environment (Google Glass [13]), medical visualization for surgeon's natural view of the operation [14], and military simulation and training

Fig. 1. A single ERP signal is elicited by a specific sensory and cognitive event. ERP is unique for individuals that different people will have distinct response with the same stimulus.

for either dangerous or costly situation [15]. According to the analysts [16], [17], the HMD market is expected to reach up to USD 15.25 billion by 2020.

2.1.2 Authentication Challenges

As the market prospects grow substantially, authentication issues for smart headwear have also drawn wide attention [18]. For all personal devices, secure authentication is vital as they store an enormous amount of information about user's privacy. Particularly, in wearable devices, the information is even more privacy-sensitive than that of mobile phones. For instance, an optical HMD with a built-in camera, like Google Glass, can record everything that a user is staring. If an unauthorized access is allowed, this could result in a leakage of user's financial and health information.

To date, existing authentication approached for HMD are limited in multiple aspects. Since HMDs are lack of either physical keyboards or touchscreen, current authentication systems in HDM often rely on additional mobile devices, which must be carried along, registered, and paired via a wireless connection (e.g., Bluetooth). For hands-off devices, this authentication mechanism is not only inconvenient but also vulnerable to hacking if the paired device is lost or stolen. Moreover, Bluetooth connection with the smartphone could be another critical security flaw [19]. In fact, modern technological advance provides better security mechanisms using biometrics, such as eye blinking [20], head movement [21], and hand gesture [22] for authentication in HMD. Yet, addressed methods are not perfectly trustworthy because a majority of biometrics can be surreptitiously duplicated or adversely revealed by attackers [23].

2.2 Brain Response to Visual Stimuli

2.2.1 ERP Rationale

ERP is a stereotyped brainwave response to a specific sensory, cognitive, or motor event. Part of what makes each human unique is their memory. No two people have had exactly the same experiences. Importantly, no two people interpret the similar events in exactly the same way. Each person's interpretation of an event is based on their semantic memory, a part of memory that includes a person's knowledge about what images depict and how they relate to own experiences [24]. Thus, semantic memory is individually unique in this way, and the activity of semantic memory is visible in the scalp-recorded ERP, as shown in Fig. 1.

2.2.2 Characteristics of ERP

Cancelable. In traditional authentication systems, users can easily replace the password when their credential is divulged. As an analogy to this, we argue that hard-cancellation can be achieved with ERP biometrics by changing visual stimuli. No person has exactly the same experience and memory on different events. Since the ERP is a stereotyped response to a particular event, we claim that the change of the event can alter the characteristics (e.g., shape, occurrence duration) of individual's ERP signal and provide new ERP signatures for the password reset.

Stable. Electroencephalogram (EEG) is a type of brainwaves that is often collected without stimulation. Therefore, the performance of EEG biometrics could be highly unstable as it depends on individual's emotional and physical states at the moment of authentication. Moreover, Ruiz-Blondet et al. [25] demonstrated typical EEG signals cannot reflect narrow, specific and cognitive processes as they are not captured time-locked to any stimulus. In our study, we present much more stable authentication method by utilizing the ERP signal, a stimulus-averaged signal that is time-locked to a specific event.

Non-Volitional. In the absence of stimulation, EEG can be volitionally modulated. For instance, a volitional control of neural activities can be achieved by real and imagined movements and cognitive imagery [26]. Thus, without stimulation, EEG can be controlled by conscious thinking of the user, which denotes that EEG is less secure to be used for authentication in case that users intentionally disclose their EEG credentials. In contrast, ERP biometrics are evoked by the stimulus, and therefore it is not under control of the user. This characteristic prevents the user from manipulating the brainwave contents purposely [25].

3 ERP Authentication Framework

3.1 Framework Overview

Our proposed system comprises three modules: visual stimuli selection, ERP signal acquisition, and signal pattern analysis. Primarily, a series of stimuli are selected according to our designated stimuli selection strategy. Brainwave signals are then acquired and averaged into the stimulus-averaged ERP signal. Then, the ERP signals are filtered, and the features are extracted via autoregressive model (AR), power spectral density (PSD), and eigenvector. Lastly, the classification of feature vectors is performed via support vector machine. The illustration of the ERP-based authentication system is shown in Fig. 2.

3.2 Visual Stimuli Design

Design Fundamental. To generate effective ERP biosignals, we use a distinct stimulation protocol that consists of a large set of various stimuli. As an analogy to a strong personal identification number (PIN) that requires a mix of numbers, letters, and special characters, (e.g., 1E@2R!3P), our brain password design also includes a mixture of various visual stimuli to enhance the "brain password" strength.

The criterion of stimuli selection is that the chosen stimuli must stimulate certain brain areas and reflect certain functional capabilities of the human brain. In this way, our brain password can satisfy the design diversity, thus form a secure and robust credential. As shown in Fig. 3, three special areas are existing at the back of the human brain, including intraparietal sulcus, inferior parietal lobule, and temporo parietal junction, each of which corresponds to the

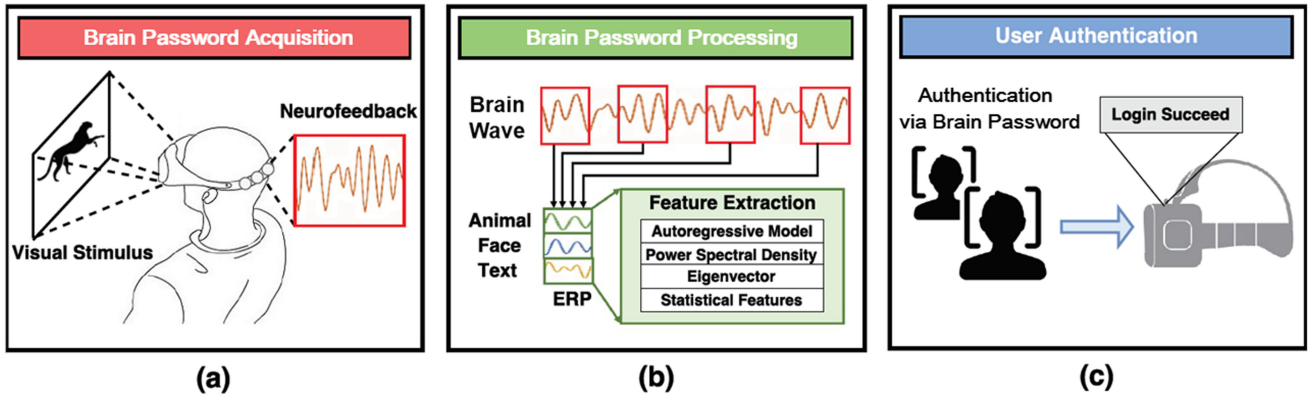


Fig. 2. An ERP-based HMD authentication system framework is illustrated. (a) When a user attempts to access a head-mounted device (such as a VR headset or Google Glass), a set of visual stimuli will be displayed on the display optic, and the dry sensors implemented in the device will measure brain-responses. (b) Obtained brain signals will be processed and analyzed. (c) The ownership will be identified by comparing with pre-stored templates of the device owner.

dedicated function of human brain. Specifically, intraparietal sulcus controls the declarative memory [27], inferior parietal lobule processes the face recognition [28], and temporo parietal junction manages the reading comprehension [29]. When a certain function is evoked, a distinct characteristic of the brain waveform is exhibited. In our design, pictures of *animal, celebrity human face, and the segment of texts* are selected as the effective stimuli for aforementioned brain areas to process declarative memory, face recognition, and reading comprehension, respectively. The examples of three visual stimuli are shown in Fig. 4.

The rationale for choosing pictures of animal for the declarative memory is that one's semantic memory on the appearance of a certain animal is highly individualized [30]. For example, a person who has suffered a spider bite will react differently to a spider picture than the person who has never been suffered from the spider. Moreover, the brain activation of the people with particular emotion to a certain category of animal is different from the brain activation of the people who don't possess such emotional state when the visual representation of that category of animal is exposed [30]. As for the human face, neurophysiology studies [31], [32] prove that the unique subject-specific brain signals can be obtained during the human face recognition. For instance, face stimuli elicit a larger peak of the negative brain potential at 170 ms (N170) compared to the ERP evoked by non-face stimuli [33]. Furthermore, texts are used to elicit the semantic memory as it is extremely unlikely for any two people to have same ability to comprehend text. Also, texts are known to elicit a distinctive negative brain potential for each individual [34].

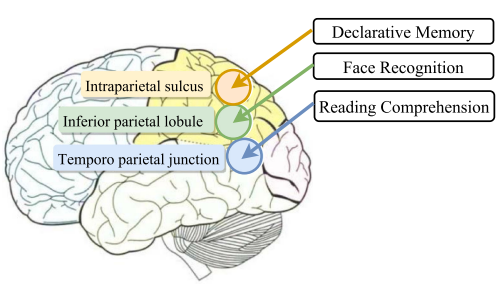


Fig. 3. Certain areas of human brain largely influence certain cognitive functions.

Visual Stimuli Selection. To choose effective images from three stimuli types, we require the ERP signal from each type of stimuli to be distinct from the ones from other types of stimuli, such that each ERP signal can significantly reflect the attributes of their corresponding brain areas. Therefore, we aim at selecting stimuli whose ERP signals can achieve maximization of the dissimilarity among them. Specifically, let p(t) be the continuous-time 2D ERP signal and T_s be the sampling period. The discrete ERP sample for each stimulus can be written as:

$$p_i = p(iT_s), \tag{1}$$

For the jth ERP signal from animals stimuli, it can be written as:

$$\mathbf{a}_j = \left\{p_1^{a,j}, p_2^{a,j}, \ldots, p_{N_s}^{a,j}\right\}^T, j = 1, 2, \ldots, N, \tag{2}$$

where N_s denotes the number of the sample size in the ERP signal, and N denotes the total number of the ERP for each type in the pool of collected data. The superscript a indicates that the signal belongs to the animals stimuli category. Likewise, the ERP signal from texts and celebrity human faces can also be written as:

$$\mathbf{t}_j = \left\{p_1^{t,j}, p_2^{t,j}, \ldots, p_{N_s}^{t,j}\right\}^T, j = 1, 2, \ldots, N, \tag{3}$$

$$\mathbf{f}_j = \left\{p_1^{f,j}, p_2^{f,j}, \ldots, p_{N_s}^{f,j}\right\}^T, j = 1, 2, \ldots, N. \tag{4}$$

The superscripts, t and f, denote the signal belonging to the texts and faces stimuli category, respectively.

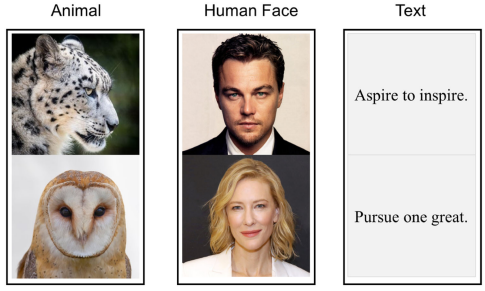


Fig. 4. Examples of visual stimuli, including animals, celebrity human faces, and texts.

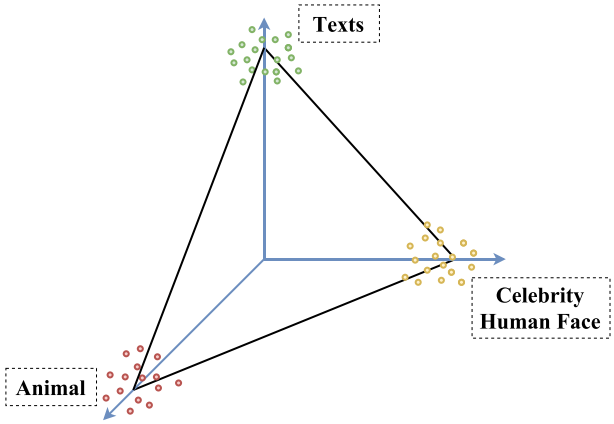


Fig. 5. A geometric illustration on visual stimuli selection. Images of animals, celebrity human faces, and texts are distributed in the 3D space as three clusters. We aim to find three dots from three clusters, respectively, that expand the triangle with the maximum perimeter.

ERP signals corresponding to the same stimulus can be expressed and mapped as a dot in a high-dimensional space, where each point has the dimensionality of N_s. For ease of representation, we depict the geometric relationship of ERP signals in a 3D space, as depicted in Fig. 5. The ERP signals from the same type of stimuli are aggregated as a set, namely, $\mathbf{A}$ for animals, $\mathbf{T}$ for texts, and $\mathbf{F}$ for celebrity faces. To maximize the diversity among the ERP signals from different types, we aim to find a triangle, as shown in Fig. 5, which has the largest perimeter. Thus, the visual stimuli selection can be formulated as follows:

$$\underset{i,j,k}{\text{maximize}} \left\|\mathbf{a}_i - \mathbf{t}_j\right\|_2 + \left\|\mathbf{a}_i - \mathbf{f}_k\right\|_2 + \left\|\mathbf{t}_j - \mathbf{f}_k\right\|_2, \tag{5}$$

$$\text{s.t.} \;\; \mathbf{a}_i \in \mathbf{A}, \mathbf{t}_j \in \mathbf{T}, \mathbf{f}_k \in \mathbf{F}, i, j, k = 1, 2, \ldots, N. \tag{6}$$

By solving the above formulation, we can use the solution set $\{i, j, k\}$ as the ERP stimuli set.

Password Set Expansion. We can define the size of the ERP stimuli set by finding the sub-optimal solution with a certain dimension in Eqs. (5) and (6). This is similar to expanding the PIN password length from "1@a" to "1@a2!b". In this study, we define the size of the ERP password set as N_p, where we consider one combination of three stimuli types (one triangle) as one password set ($N_p = 1$). The performance of various N_p values are evaluated and discussed in Section 7.4.7.

4 System Implementation

4.1 System Overview

Fig. 6 shows the flowchart of our proposed system. A set of visual stimuli is selected from the database and displayed to the user through the VR headset. The generated ERP signal is extracted and analyzed for the later matching with the owner record. If they matches, the user is considered as the owner. Otherwise, she is rejected as the intruder.

4.2 ERP Acquisition Device

To capture the ERP data, our team has developed an ERP brain sensor headset, which is equipped with dry electrodes. Such electrodes utilize a set of angled legs and permits the legs to flex outward under pressure which helps push aside hair for better contact. The sensors are coated with metallized paint for conductivity providing low impedance contact (100-500 kΩ) to suppress noise in the ERP acquisition. The headset employs the channel P3, Pz, and P4 (International $10-20$ System) with two grounds (Fp1 and Fp2) and reference on A1 (See Fig. 7). The brain sensor headset can conveniently collect brainwave signals at the sampling rate of 1000 Hz. The adoption of 1000Hz sampling rate will prevent losing any useful brainwave components. First, some brainwave, such as "high gamma" activity brainwave, can reach up to approximately 200 Hz [35]. Second, EEG signal is no perfect sine wave and that it will have significant harmonic content. Thus, the 1000Hz sampling rate, that is higher than the Nyquist rate, is an appropriate choice. Then, the collected data can be saved locally or streamed to a computer via Bluetooth.

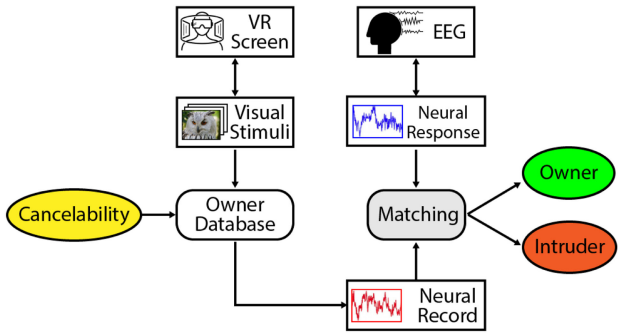


Fig. 6. The flowchart of the proposed ERP authentication system. The neural response is compared with the pre-stored neural record to determine whether the unknown user is the authentic user.

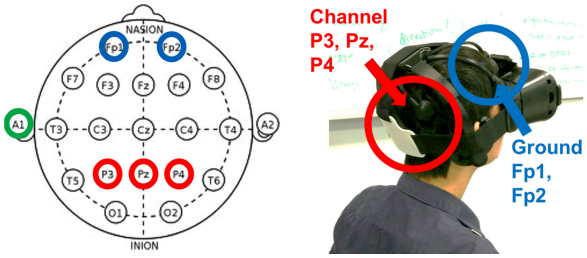
4.3 Electrode Placement

In standard practice, 32 to 64 electrodes are used for ERP measurement, and the number of electrodes sometimes raises up to 256 to obtain the detailed information [36]. However, the implementation of multiple electrodes in the HMDs is problematic due to the heavy weight, low cost-efficiency, and highly complex data acquisition process [23]. Therefore, we customized a sensory headset that is suitable for HMD applications. Our brain sensor device contains three channels (i.e., P3, Pz, and P4) on the parietal lobe.

According to previous studies [37], [38], [39], brain-computer interface (BCI) classification accuracy can be significantly increased by utilizing the parietal electrodes P7, P3, P4, Pz, and P8 because the negative peak of ERPs in the parietal region is unique compared to other regions. Also, since the parietal lobe has an important role in the recollection of episodic memory [40], the parietal electrodes are highly recommended as an alternative to using the complete EEG channel set. More importantly, as shown in Fig. 7a, P3, Pz and P4 are placed on the brain areas addressed in Section 3.2. Also, since the headband of HMD is typically placed on the back, these electrodes can be easily implemented in the headband, providing more convenient and non-invasive data acquisition process.

4.4 Motion Artifacts Suppression

Motion artifacts generated by the head movement may compromise the ERP recordings. However, it is inevitable while



(a) Standard EEG location (b) Actual positions

Fig. 7. 7(a) shows the standard electrode location in International 10-20 System. The electrode in green represents the reference, the electrodes in blue are grounds, and the electrodes in red reflect the channels used. 7(b) shows the placement of the Dry EEG Headset.

wearing smart headwears. In our proposed method, the automatic epoch rejection removes the data epoch with extreme artifact noises using measurement statistics including mean, standard deviation, skewness, kurtosis, and median. Then, infinite impulse response filter reduces high-frequency noises. To further compensate artifact noises, we applied a channel-based artifact template regression procedure and subsequent spatial filtering approach [41], which removes the ambulation-related movement artifacts.

5 ERP Processing

5.1 Pre-Processing

Pre-processing is applied to improve the resolution of brain signals. After obtaining a full EEG waveform, the signal is segmented from the start to the end of the stimulus hit. Thus, each ERP segment has a length of 200 milliseconds. Automatic epoch rejection [42] is applied at the probability threshold of 2.5 to remove the segments with abnormal electrode activity (e.g., fluctuation triggered by motion or sounds). Then, four ERP segments of the same type are averaged into a single stimulus-averaged ERP signal. These ERP signals of animal, face, and text type are combined into one vector. Therefore, there is 600 milliseconds stimulus-averaged ERP template per channel per subject. Then, an infinite impulse response (IIR) Butterworth filter is employed to produce a zero phase-shift.

5.2 Feature Extraction

Each channel has 280 feature elements, and the feature vector of the channel is attached to the feature vector of other channels. Therefore, the final length of one feature vector is 840. The following features are extracted for each feature vector:

Autoregressive Model. We utilize three 6th order autoregressive (AR) models [43] to extract ERP features. AR model is advantageous with short data segments because the frequency resolution of AR spectrum is infinite and does not depend on the length of analyzed data [44]. Since our ERP signals are short data segments, AR model is suitable for our system. By definition, the AR model is a linear difference equation in the time domain:

$$X_t = \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t, \tag{7}$$

where X_t is the signal at the sampled point t, p is the order of the model, a_i is the AR coefficient, and ε_t is an independent and identically distributed white noise input [45]. To obtain normalized autoregressive (AR) parameters, we employ the Yule-Walker method [46], which exploits the approximate of the autocorrelation data function. Then, the Burg method [47] is utilized to reduce linear prediction errors. Lastly, the covariance and modified covariance methods are used to minimize the forward and backward prediction errors. Since each model consists six parameters, 24 AR coefficients are obtained for each channel. With all three channels, there are 72 features attached to the vector.

Power Spectral Density. To accurately detect the spread of power with respect to frequency, the power spectral density (PSD) estimate is obtained by the Welch's overlapped segment averaging estimator [48]. First, ERP signals are divided into frames of 128 to utilize periodogram method for ERP application. The periodogram method is based on Fourier transform and known as non-parametric spectral estimation method. Then, the Welch power spectrum estimates the PSD by averaging modified periodograms. We extract 128 features from the estimates for each channel and consequently attach 384 features to the feature vector.

Eigenvector. Since the skin electrode interfaces in dry EEG may induce signal noises, the eigenvector spectral estimation method is used to compensate the effect of the noises. The eigenvector method is known to provide a suitable resolution for artifact corrupted signals by calculating a pseudo-spectrum estimation, which is defined as [44], [49]:

$$P(f) = \frac{1}{\sum_{j=i+1}^{N} \left| V_j^H e(f) \right|^2 / \lambda_j}, \tag{8}$$

where $V_j^H e(f)$ represents a Fourier transform, N is the dimension of the eigenvectors, i indicates the integer value of the dimension of the signal subspace, and λ_j represents the eigenvalue of the matrix. ERP signals are divided into frames of 128, and the pseudo-spectrum is measured by estimates of the eigenvectors. We extract 128 features for each channel, and total 384 features are obtained for the feature vector.

5.3 User Authentication

The user authentication process is described as below. Initially, owner's template is stored in the system. Then, the anonymous user attempts to access the system by wearing the smart headwear device. After detecting the user presence, the system provides a series of stimulus and elicits brain signals of the unknown user. The stimulus-averaged ERP signal from the corresponding user is then verified based on the pre-stored templates. During the authentication process, we employ support vector machine (SVM) with a radial basis function (RBF) kernel [50] for the classifier. The choice of the classifier will be further discussed in Section 7.4.5. SVM with RBF kernel enables classification operation in a high-dimensional, implicit feature space without ever computing the coordinates of the data in the input space, where two parameters γ and C dominates the kernel function. γ can be seen as the inverse of the radius of influence of samples selected by the model as support vectors and C trades off misclassification of training examples against simplicity of the decision surface. In our study, γ and C of RBF function are chosen as 0.001 and 10000, respectively.

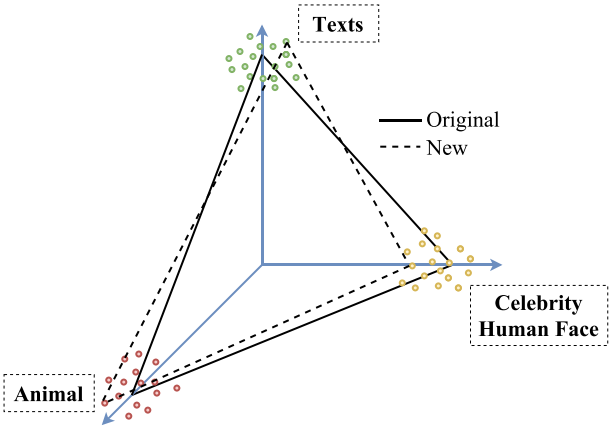


Fig. 8. Illustration of stimuli update strategy, where the original password design is depicted using real line and the new one is depicted using dash line. The new triangle should also comprise of dots from those three clusters, while it should also be far away from the original one.

6 Cancelability Against Attack

Traditionally, once a human biometric, such as iris or fingerprint, is divulged, the authentication system is compromised and no longer safe to use. Comparing with these biometrics, ERP-based brain password is superior because the originally stored credential of brainwave can be canceled if divulged. In other words, our system updates the in-use stimuli to avoid any potential risk. In practice, when a user need to change their password, the system will present a large number of images from the pool to the user and record the brainwave signal, then there is an offline phase where a new password is chosen corresponding to a subset of the images where the selection of that subset follows a stimuli update strategy. In this section, we will deliberate the stimuli update strategy to cancel ERP credentials.

6.1 Stimuli Update Strategy

The update strategy is illustrated in Fig. 8, where the original password design is depicted using real line and the new one is depicted using dash line. The candidates for new visual stimuli must satisfy two conditions:

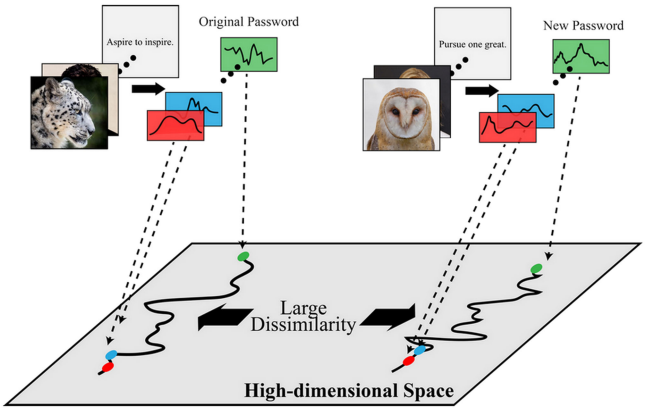


Fig. 9. Illustration of maximizing the dissimilarity in the joint spatio-temporal domain. The original password design is depicted on the left, and the new one is depicted on the right. The update strategy intends to maximize the difference (i.e., the designated distance) between the original ERP-based biometric credential and the newly generated one.

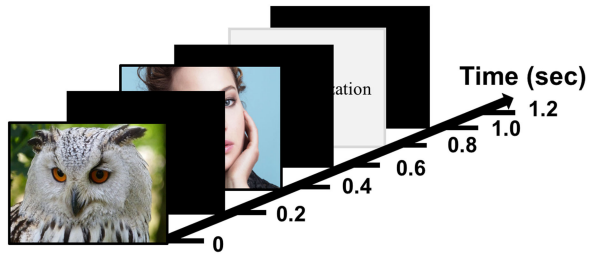


Fig. 10. The time interval between images. Each image flashes for 200 milliseconds, and it takes 200 millisecond to switch to the next image. This image sequence is shown for four times, and four brain responses from each image are combined into an aggregated ERP response.

1) the new brain password should achieve comparable authentication performance comparing with the original one. Therefore, the new stimuli should also comprise of images from the three diverse categories separately.
2) the ERP signals evoked by these images should be distinct from the ones evoked by the original images, which is analogical to the case where we are not allowed to use the previously used passwords when resetting passwords. In this way, we guarantee the two passwords are disparate enough that the original brain password is not accessible to the system configured with the new brain password. In other words, we aim to maintain an extremely low false acceptance rate by preventing unauthorized access.

As users viewing images sequentially, visual stimuli and the corresponding ERP signals can be considered as time series signals, which will be later described in Section 7.2 and shown in Fig. 10. In the meantime, for a specific image, its ERP signal can be expressed as a dot in high-dimensional space. Therefore, the time series of ERP signals exhibit spatio-temporal attribute. To quantify the dissimilarity of ERP signals that are generated by the original and new selection of images, we propose a dissimilarity metric, i.e., spatio-temporal warping distance, and compare the two password designs (i.e., ERP stimuli sets) in the joint spatio-temporal domain. Our goal is to find the maximum dissimilarity between two password design in terms of the spatio-temporal warping distance. The concept of maximizing the dissimilarity in the joint spatio-temporal domain in depicted in Fig. 9.

6.2 Dissimilarity Measurement Metric

In the following, we will elaborate the design of spatio-temporal warping distance as the dissimilarity measurement metric.

Spatial Domain Analysis. suppose the jth images are considered for both original and new ERP signals, and the ERP signals can be represented in the form of vectors, as follows:

$$\boldsymbol{\gamma}_j = \left\{\mathrm{a}_j^T, \mathrm{t}_j^T, \mathrm{f}_j^T\right\}^T = \left\{p_1^{\gamma,j}, p_2^{\gamma,j}, \ldots, p_{3N_s}^{\gamma,j}\right\}^T, j = 1, \ldots, N, \quad (9)$$

$$\widetilde{\gamma}_j = \left\{\widetilde{\mathrm{a}}_j^T, \widetilde{\mathrm{t}}_j^T, \widetilde{\mathrm{f}}_j^T\right\}^T = \left\{\widetilde{p}_1^{\gamma,j}, \widetilde{p}_2^{\gamma,j}, \ldots, \widetilde{p}_{3N_s}^{\gamma,j}\right\}^T, j = 1, \ldots, N. \quad (10)$$

where each element is as defined in Eqs. (2), (3), and (4), and the superscript γ indicates the element belongs to γ. Both $\boldsymbol{\gamma}_j$ and $\widetilde{\gamma}_j$ have the dimension of $3N_s$.

For the pair of $\boldsymbol{\gamma}_j$ and $\widetilde{\gamma}_j$, each element in the vector is normalized by dividing the sum of all elements in the vector, written as:

$$q_i^j = \frac{p_i^{\gamma,j}}{\sum_{i=1}^{3N_s} p_i^{\gamma,j}},\ \widetilde{q}_k^j = \frac{\widetilde{p}_k^{\gamma,j}}{\sum_{k=1}^{3N_s} \widetilde{p}_k^{\gamma,j}}. \tag{11}$$

Here, we use q_i^j and $\widetilde{q}_k^j$ to denote the normalized value, and the superscript γ is removed since there is no ambiguity for the symbol q and $\widetilde{q}$. Then we define the cost c_{ik} of transporting between ith data from γ_j, which is q_i^j, and kth data from $\widetilde{\boldsymbol{\gamma}}_j$, which is $\widetilde{q}_k^j$. Specifically, we use the euclidean norm for the cost definition.

The next task is to find a flow, $\mathbf{F}(i,k) = f_{ik}$, such that the matching work between two datasets γ_j and $\widetilde{\boldsymbol{\gamma}}_j$ will have the least cost [51]:

$$\text{minimize} \sum_{i=1}^{3N_s}\sum_{k=1}^{3N_s} c_{ik} f_{ik}, \tag{12}$$

$$\text{s.t.} \sum_{i=1}^{3N_s} q_i^j = \sum_{k=1}^{3N_s} \widetilde{q}_k^j, \tag{13}$$

$$f_{ik} \geq 0, 1 \leq i \leq 3N_s, 1 \leq k \leq 3N_s, \tag{14}$$

$$\sum_{k=1}^{3N_s} f_{ik} \leq q_i^j, 1 \leq i \leq 3N_s, \tag{15}$$

$$\sum_{i=1}^{3N_s} f_{ik} \leq \widetilde{q}_k^j, 1 \leq k \leq 3N_s, \tag{16}$$

$$\sum_{i=1}^{3N_s}\sum_{k=1}^{3N_s} f_{ik} = \min\left(\sum_{i=1}^{3N_s} q_i^j, \sum_{k=1}^{3N_s} \widetilde{q}_k^j\right). \tag{17}$$

Once the above problem is solved, and we have found the optimal flow $\mathbf{F}$, the spatial matching (SM) metric is found as the matching work normalized by the total flow:

$$\text{SM}(\widetilde{\gamma}_j, \gamma_j) = \frac{\sum_{i=1}^{3N_s}\sum_{k=1}^{3N_s} c_{ik} f_{ik}}{\sum_{i=1}^{3N_s}\sum_{k=1}^{3N_s} f_{ik}}. \tag{18}$$

Temporal Domain Analysis. Suppose the password set size $N_p > 1$, which means there are more than one image set from three clusters, we can incorporate the temporal domain analysis in addition to the spatial domain analysis. To measure the similarity between these two sequences of images that illustrated in Fig. 10, an $N_p \times N_p$ matrix $\mathbf{D}$ is created, called *distance matrix*. The value of the (m^{th}, n^{th}) element in $\mathbf{D}$ represents the distance $d(\widetilde{\gamma_{\mathbf{n}}}, \gamma_{\mathbf{m}})$ between two sets of ERP signals $\widetilde{\gamma_{\mathbf{n}}}$ and $\gamma_{\mathbf{m}}$. Then the SM defined in Eq. (18) is adopted as the distance metric, and we can obtain:

$$\mathbf{D}(n,m) = d(\widetilde{\gamma}_n, \boldsymbol{\gamma}_m) = \text{SM}(\widetilde{\gamma_n}, \gamma_m). \tag{19}$$

With the guidance of the distance matrix, the shortest warped path through the matrix can be derived [52]:

$$cd(n,m) = \text{SM}(\widetilde{\gamma_n}, \gamma_m) + \min\begin{cases} cd(n, m-1) \\ cd(n-1, m) \\ cd(n-1, m-1) \end{cases}, \quad 1 \leq n \leq N_p, 1 \leq m \leq N_p. \tag{20}$$

where $cd(n,m)$ is the current minimum cumulative distance for $\mathbf{D}(n,m)$, and the initial setting is $cd(0,0) = 0, cd(0,m) = cd(n,0) = \infty$.

After that, the overall minimized cumulative distance $cd(N_p, N_p)$ can be found. Finally, the spatio-temporal warping distance is calculated as:

$$Dist = cd(N_p, N_p). \tag{21}$$

Overall, our aim is to find a new design that has the maximum $Dist$ to the original design.

7 Performance Evaluation

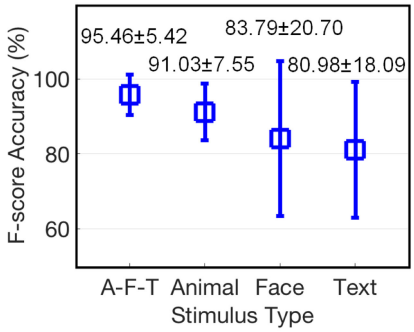
7.1 Participants

In the pilot study, ERP signals are obtained from total 179 adult participants with a mean age of 29.85 and standard deviation of 7.72. These participants are comprised of undergraduate/graduate students and faculties/staffs from university, students from high school, and volunteers from neighborhood with various occupations, all of who had no brain or ocular diseases. Among 179 participants, 93 of them are male participants, and 86 of them are female participants. Consent forms for participation in the research study were obtained at the time of the study, and all participants have received a comprehensive description of the experimental procedures. As mentioned above, electroencephalography is a safe monitoring method with no side effects [53]. Moreover, our headset is in dry form that does not require gel or other fluids. To alleviate possible eye irritation that may occur due to the various stimuli used in the procedure, we avoided the use of extremely bright colors and flashing lights.

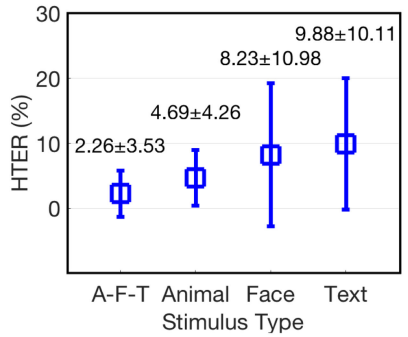
As described above, the system evaluation relies on a strategically developed experiment that will involve a cohort of participants. We hold an existing active IRB protocol that allows for recording brainwave from adult human participants for user authentication. All the evaluations tightly follow the rule of IRB regulation.

7.2 ERP Acquisition

The general procedure for ERP collection was as follows. The ERP data collection was conducted in a normal lab office environment with ambient sounds and noise. People in the lab office other than participants have no obligation to keep quiet, instead, they are free to act as normal including talking, walking, typing, closing doors, or moving chairs, etc. Participants were seated on a comfortable chair wearing the HMD while ERPs were collected. They were instructed to pay attention to the presented images without thinking about anything else. In our ERP acquisition protocol, three types of images are displayed on the screen in a certain order. The order of the stimulus presentation is from Animal, human Face to Text (short for A-F-T). When this stimuli sequence with certain images repeats for four times, the acquired EEG signal undergoes the ERP processing method (see Section 5) and produces a single stimulus-averaged ERP, which we simply refer to an ERP signal, for each stimulus type. Each image is flashed for only 200 ms to avoid the use of exploratory eye movements, and 200 ms interval is applied in between two images to make each stimulus independent of the previous stimulus



(a) F-score comparison among combined and separate stimuli types. A-F-T achieves the best accuracy of 95.46% with the least standard deviation of 5.42% comparing with separate stimulus type.



(b) HTER comparison among combined and separate stimuli types. A-F-T achieves the lowest error rate of 2.26% with the least standard deviation of 3.53% comparing with separate stimulus type.

Fig. 11. Performance comparison among different stimulus type. A-F-T indicates the combination of animal, face, and text stimulus type.

(see Fig. 10). In our experimental protocol, the acquisition of ERPs for the animal, human face, and text took approximately 4.8 seconds. The appropriate duration of stimulus presentation (equivalent to number of images) is further investigated in Section 7.4.7.

7.3 Experiment Description

The data are collected in two sessions. The data from the first session is used to evaluate the system performance and cancelability, and the data from the second session is used for a longitudinal study. Among 179 participants, 80 have participated in the second session. Because some data from 2 participants are damaged, the valid participants for the longitudinal study is 78 with the average age of 27.36.

As there are total 179 participants, one of the subjects acts as an owner once while the remaining subjects act as attackers. This process repeats for all subjects. Here, 10-fold cross validation is used to prevent overfitting. The data set is randomly separated into 10 equal-sized subsets. For each trial, one of the 10 subsets is used as a test set, and remaining subsets are used as a training set. This cross-validation is repeated with each of the subsets.

For each session, the data collection task is organized in a series of 300 images with 100 images for each stimulus type. As mentioned in Section 7.2, a series of same images repeats for four times. Thus, there are 25 different images among 100 images for each type. In other words, the number of stimulus-averaged ERP (N) in the pool of each animal, human face, and text set is 25, which corresponds to the total number of dots in each cluster. For the authentication, one dot for every cluster (one triangle) is used for one-set password ($N_p = 1$), two dots for every cluster (two triangles) are used for two-set password ($N_p = 2$), and three dots (three triangles) for every cluster are used for three-set password ($N_p = 3$). The maximum number of set is N, which is equivalent to 25. We used the one-set password for all evaluations except for Section 7.4.7. To produce multiple ERP templates, we repeat the data collection task 20 times for each participant.

7.4 System Performance

7.4.1 F-Score Accuracy

The accuracy (ACC) [54] is predominantly used for the statistical classification. However, ACC is an inappropriate accuracy metrics when negative and positive classes are not balanced. Thus, to avoid an unbalanced accuracy measurement, we evaluate our system performance based on f-score accuracy (F_1), which is also known as a harmonic mean of precision and recall. Particularly, f-score accuracy provides outstanding performance with high negatives and low positives. Considering the intrinsic unbalance of positive and negative samples in the biometric security study, f-score accuracy is preferred for the sake of non-sensitivity to class imbalance.

Mathematically, F_1 is defined as follows:

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad [\%]. \tag{22}$$

where TP is an abbreviation of true positive, FP is a false positive, and FN represents a false negative.

Figure 11a depicts the f-score comparison among various stimulus types. As shown, A-F-T indicates the combination of animal, face, and text stimuli that is designed based on our visual stimuli model (see Section 3.2). The stimuli for animal, face, and text types are identical to the pictures used in A-F-T. Among four types, A-F-T achieves the best accuracy of 95.46 percent with the least standard deviation (STD) of 5.42 percent. The accuracy of A-F-T is higher than that of animal, face, and text stimuli by 4.43, 11.67, and 14.48 percent, respectively. Moreover, the STD of A-F-T is lower than other three types by 2.13, 15.28, and 12.67 percent, respectively. The results prove that our visual stimuli model improves security and robustness of the brain password by satisfying the design diversity.

We compared our system with the widely used commercial fingerprint scanning system. During the daily fingerprint authentication in mobile devices, usually only partial area of the fingerprint will be scanned by the sensor [55]. According to references [56], [57], partial coverage area of 50 percent can achieve around 96 percent accuracy tested on both FVC2002 DB2 [58] and NIST SD30 [59] database. While our system achieves 95.46 percent average accuracy, which is comparable to the performance of partial fingerprint scanning, and hence applicable in practice.

7.4.2 Half Total Error Rate

Half total error rate (HTER) is a metric method that is widely used in brain biometric literature [60], [61]. It

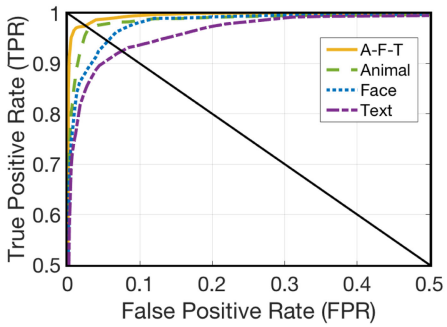


Fig. 12. The average receiver operating characteristic curve. The equal error rate, a rate that corresponds to an equal probability of true positive rate and false positive rate, is the x-value of intersection point between the curve and the diagonal of the unit square.

measures the detection performance by averaging the false rejection rate and the false acceptance rate:

$$HTER = \frac{FAR + FRR}{2} \quad [\%], \tag{23}$$

where FAR refers to the false accept rate, an ability of a non-authorized user to access a system, and FRR is the false rejection rate that occurs when a user is not matched to own biometrics profile [62].

In brief, our A-F-T has the average FRR of 4.30 percent with the average FAR value of 0.22 percent. The average HTER value is 2.26 percent, and its STD is 3.53 percent. FAR value is lower than FRR value because our authentication scenario contains more log-in attempts of the attacks than the owner's attempts. In Fig. 11b, the HTER of four stimuli types are illustrated. The performance of A-F-T outperforms other three types as A-F-T has lower HTER than the animal by 2.43 percent, face by 5.97 percent, and the text by 7.62 percent. These HTER results conform to the previous f-score comparison in Fig. 11a. Moreover, the STD of A-F-T is lower than the STD of animal, face, and text by 0.73 percent, 7.45 percent, and 6.58 percent, respectively. A relative small HTER and STD of A-F-T prove that our visual stimuli design improves the universality of brain biometrics upon 20 subjects.

7.4.3 Receiver Operating Characteristic Curve

For a comprehensive evaluation of the system performance, a receiver operating characteristic curve (ROC) is investigated. By definition, it visualizes the sensitivity or TPR (true positive rate) against FPR (false positive rate) as the threshold is varied. As the curve follows the top-left portion of the graph, the system has a high sensitivity and specificity and is more accurate. In Fig. 12, the average ROC curve of A-F-T, animal, face, and text stimulus type are plotted. Among four curves, A-F-T follows the most upper-left portion of the graph, indicating that our system is robust and feasible.

7.4.4 Equal Error Rate

The equal error rate (EER), a rate that corresponds to an equal probability of an acceptance error and rejection error, can be derived from the average ROC curve. Specifically, the x-axis value of intersection point between the curve and the diagonal of the unit square is known as EER. More specifically, the EER value of A-F-T is $2.503 \pm 0.05\%$ and the EER of animal, face, and text are $3.114 \pm 0.06\%$, $5.559 \pm 0.08\%$, and $7.517 \pm 0.1\%$ with a 95 percent confidence interval [63], respectively (derived from Fig. 12). Again, A-F-T achieves lowest EER, which indicates that our visual stimuli model increases the system performance.

TABLE 1
The Different Classifiers Comparison

	bagged trees	linear	polynomial	RBF
F1 (%)	90.77	89.84	93.18	95.46
HTER (%)	8.254	9.36	4.93	2.26

7.4.5 Classifier Impact

We compared four different classification techniques to select the best classifier for our application, including support vector machine (SVM) with a linear kernel, a polynomial kernel, a radial basis function (RBF) kernel, and the bootstrap aggregated (bagged) trees. Bagged trees is a classification method that improves the predicative accuracy by creating multiple versions of a predictor and using these predictors to obtain an aggregated predictor [64], [65]. Parameters of each classifier are tuned to achieve the best performance. γ and C of RBF function are 0.001 and 10000, respectively. For bagged trees, maximum number of splits is 20, and the number of learners is 30. The F1 and HTER results for A-F-T are shown in Table 1. The SVM with RBF kernel showed the best performance, which is adopted for the classification.

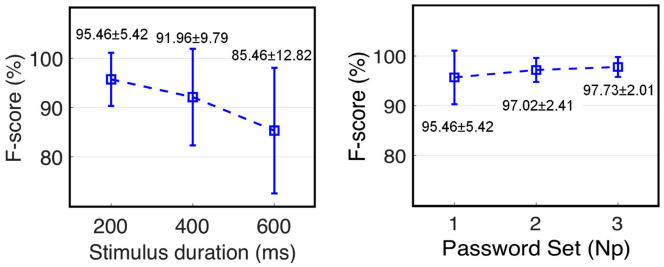
7.4.6 One-Class Classification

In our user authentication described in Section 5.3, binary-class classification is adopted by taking the training samples from would-be attackers as well as the ones from the authentic user. To further prove the robustness of our approach, we model the authentication as a one-class classification that training the classifier with the patterns of just the legal user and testing it with the data from all the would-be impostors. One-class SVM, specifically, the support vector data description approach [66] is adopted for the classification by using the LibSVM tool [67]. The F1 and HTER results for A-F-T are 93.46 percent with STD of 6.27 and 4.18 percent with STD of 4.85 percent, respectively. The results show our system has a reasonable performance with the one-class classifier.

7.4.7 Optimization of Authentication Time Efficiency

Since our authentication system targets for the smart headwear application, the optimization of the authentication time is essential. Thus, we examine several methods to optimize the authentication time efficiency.

Stimulus Duration. During the experiment, each stimulus is presented for 200 ms, and the black screen is displayed for 200 ms to separate each stimulus. By discovering the optimal stimulus duration, the authentication time can be reduced. As shown in Fig. 13a, the accuracy declines by 3.5 percent and the STD increases by 4.37 percent as the stimulus duration increases from 200 ms to 400 ms. Similarly, when the



(a) Impact of stimulus duration. The 200 ms duration setting achieves the best F-score of 95.46% with standard deviation of 5.42. Larger duration will have lower F-score and larger standard deviation.

(b) Impact of password length. Longer password set will have higher F-score. With the simplest one-set of password, we achieves 95.46% F-score.

Fig. 13. The authentication time optimization via stimulus duration and password length adjustment.

duration of stimulus exceeds 600 ms, the accuracy reaches 85.46 percent, which is 10 percent lower than the accuracy of 200 ms. Also, the STD increases by 7.4 percent at 600 ms. The reason for this phenomenon is because stimulus presented for more than 200 ms will induce exploratory eye movements, which in some extent will compromise the collected EEG signal dedicated as a response to the visual stimuli. Although the accuracy increases with decreasing duration, the stimulus duration less than 200 ms is too instantaneous for the average human reaction time to the onset of a visual stimulus [68]. Thus, the optimal stimulus duration is 200 ms.

Password Set. As described in Section 3.2, we can optimize the authentication time efficiency by adjusting the size of password set (see Fig. 13b). For one-set password ($N_p = 1$), the system accuracy reaches 95.46 percent. When two-set password ($N_p = 2$) is used, the accuracy increases by 1.56 percent and the STD decreases by 3.01 percent. When three-set password ($N_p = 3$) is employed, the accuracy is increased by 0.71 percent and the STD is reduced by 0.4 percent. This result indicates that the accuracy and stability of the system increase as the size of password increases.

Time Efficiency. As the stimulus duration and size of password set increase, the authentication time increases as well. In brief, the optimal time can be calculated as:

$$Time\ (s) = N_p \cdot N_{avg} \cdot 3\ (stimulus\ duration + 0.20), \quad (24)$$

where N_p indicates the size of password set, and N_{avg} represents the number of the segments that are averaged into a stimulus-averaged ERP, which is 4. In the formula, the interval duration (0.20 second) and the number of stimulus type (3 for animal, face, and text) are included. With the optimal stimulus duration (200 ms), one-set password takes approximately 4.80 seconds, two-set password takes 9.60 seconds, and three-set password takes 14.4 seconds for the authentication time. Also, more computation is necessary for higher N_p value. Since the authentication for smart headwear devices must be reasonably fast, we select the one-set password ($N_p = 1$) and the optimized time is 4.80 seconds.

8 Cancelability Analysis

In order to properly revoke and reissue the credential, the cancelability must satisfy two properties: *revocability* and *unlinkability* [69]. First, when a biometric database is breached, the user should be able to revoke old credential and reissue new credential derived from the same physiological trait. In other words, the system should accept new brain password while rejecting the original password without degrading the authentication performance. Second, the adversary should not be able to link or cross-match old credentials to newly generated credential. Thus, the correlation between old and new biometric instance must be low. To prove that our ERP biometric is truly cancelable, we evaluate the cancelability based on the aforementioned two properties.

TABLE 2
Performance Table for Each Stimuli Set

Trial	Recall (%)	Precision (%)	F-score (%)
Original ERP	95.68 ± 6.89	95.91 ± 4.91	95.46 ± 5.42
New ERP	94.64 ± 6.03	95.62 ± 5.11	94.87 ± 3.69

8.1 Revocability

Objectives. In this section, we verify the revocability of ERP in two ways. First, we demonstrate that new ERP generated according to our stimuli update strategy has a high accuracy to serve as a new brain password. Second, we prove that new ERP is distinguished from the original ERP, thereby corroborating its robustness against the attack using the original password.

Experiment Descriptions. The updated stimuli set is given to the participants, and 20 new ERP templates are obtained per subject. Again, each subject acts as an owner and the rest act as an attacker. To demonstrate the effectiveness of newly generated ERP, we compare the recall, precision, and f-score of the updated stimuli set to those of the original stimuli set. For the second objective, we assume the following scenario. Once user's original credential is counterfeited, the user generates new ERPs according to the stimuli update strategy and updates the user profile. The attacker uses the replication of user's original ERP to access the system configured with the new ERP. For evaluation, we randomly select a portion of new ERPs to create the updated profile and test the performance by authenticating with the remaining new ERP templates and original ERP templates from Section 7.4. We employ SVM with a 10-fold cross-validation. This procedure repeats for each subject, and the FRR and FAR are calculated. Lastly, we average the FRR and FAR of all subjects.

Results and Discussions. The evaluation results are shown in Table 2, where it reveals that the original visual stimuli will result in true negatives when adopting them to a system configured with new stimuli. The new ERP credential obtained via stimuli update strategy provides the recall, precision, and f-score of 94.64, 95.62 and 94.87, correspondingly. The STD are 6.03, 5.11, and 3.69 percent. Although the recall, precision, and f-score of the original ERPs are slightly higher by 1.04, 0.29, and 0.59 percent, these discrepancies are not significant, and the new credential still yields high recall, precision, and f-score value. Therefore, the updated strategy does not degrade our system performance. As shown in Table 3, our second revocability task

TABLE 3
Authentication of the System Configured with the New ERP

Recall (%)	Precision (%)	FRR (%)	FAR (%)
99.20 ± 1.829	99.05 ± 2.034	0.775 ± 1.805	0.789 ± 1.775

achieves a high recall and precision value of 99.20 and 99.05 percent with low FRR and FAR of 0.775 and 0.789 percent.

Fig. 14 illustrates the visual comparison of the original and new ERP signals for each stimuli type. In the figure, a data smoothing is applied to depict a general trend of the brainwave and to minimize the noises. To smooth the pattern, we employed a moving average filter that provides the average of every 15 consecutive samples of the waveform. Data points are equally weighted and contain 1/15 of the total average. Brain responses from the original stimuli set are colored in blue, and the signals in response to the new stimuli set are colored in green. Dashed box in the graph represents the stimulus duration. In comparison to the ERPs at post-stimulus period, the signals during the stimulus hit, particularly from 0 ms to 100 ms, show a larger discrepancy. This discrepancy provides evidentiary support of our hypothesis and render our argument valid; our stimuli update strategy can stimulate distinctively different ERP signals. To put it another way, replicated original credential is unlikely to be used to access the system configured with new credential. This result validates our two hypotheses. First, the ERP biometrics are truly cancelable as the change of the visual stimulus alters the characteristics of ERP. As mentioned previously, the reason is that no one has exactly the same memory on different images. For instance, the person's memory of the spider is highly likely to be different from the memory of the dog. Hence, changing the stimulus from the spider picture to the dog image elicits new characteristics in ERP. Second, our stimuli update strategy amplifies such alteration by finding the maximum dissimilarity among ERPs in response to a larger pool of images. For example, if the dissimilarity between the ERPs from the spider image and lion image is larger than the dissimilarity between the ERPs from the spider image and dog image, we incorporate the lion image to evoke new ERP. In this way, the new ERP is truly distinct from the original ERP, and the system maintains stability after the stimuli update.

8.2 Unlinkability

Objectives. Proving unlinkability between the old and new ERPs is equivalent to demonstrating the independence of each other [69], [70]. In this section, we verify the independence between the original and new ERP by performing the correlation test with the original and new ERP features. Here, we quantify the correlation by calculating the Pearson's correlation coefficient, a measure of the degree of linear relationship between two variables. If the coefficient is close to zero, there is no discernible relationship between fluctuations of the variables.

Experiment Descriptions. We employ the original and new ERP data from Section 8.1. In this experiment, we specifically use the Pearson's correlation coefficient [71], R, which is defined by the following:

$$R_{i,j} = \frac{1}{N-1}\sum_{n=1}^{N}\frac{(a_{i_n}-\mu_{a_i})}{\sigma_{a_i}}\frac{(b_{j_n}-\mu_{b_j})}{\sigma_{b_j}}, \tag{25}$$

where a_{i_n} is a feature element of an original ERP template and b_{j_n} is a feature element for a new ERP template. μ_{a_i} and σ_{a_i} represent the mean and STD of all feature elements of the corresponding original ERP template while μ_{b_j} and σ_{b_j} signify the mean and STD of the elements of the new ERP template. Every template is composed of 840 feature elements as mentioned in Section 5.2, and thus N equals to 840. However, using common feature extraction methods results in a similar trend in all ERP templates and increases the overall correlation coefficient value. Thus, before applying the Pearson's correlation coefficient, we suppress the feature trend by normalizing each template with the mean of all templates, obtained from the corresponding stimuli set and subject, as a measure of scale.

$$Normalized(A_i) = \frac{A_i}{\frac{1}{k}\sum_{p=1}^{k} A_p};\ 1 \le i \le k, \tag{26}$$

$$Normalized(B_j) = \frac{B_j}{\frac{1}{k}\sum_{p=1}^{k} B_p};\ 1 \le j \le k, \tag{27}$$

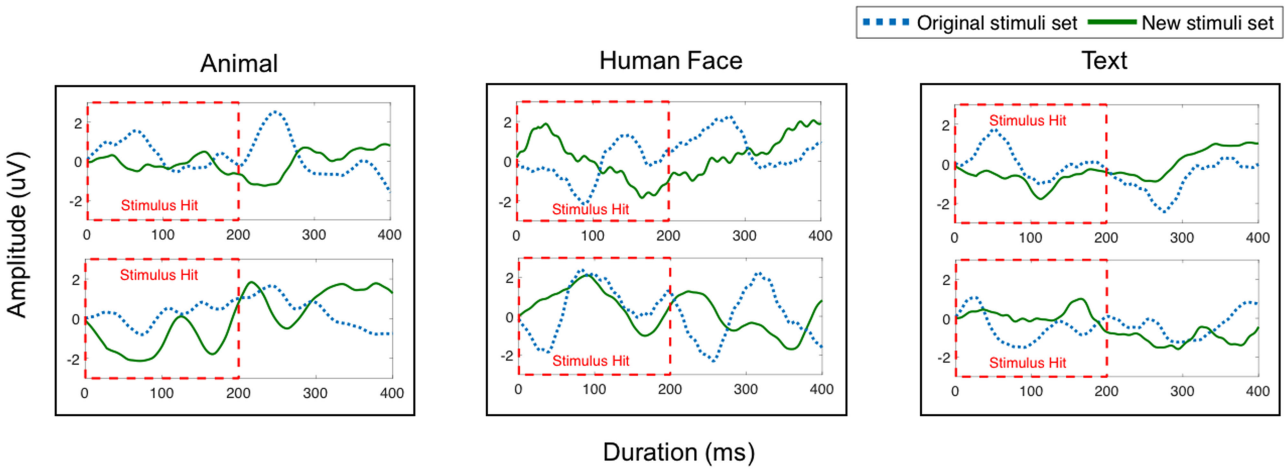
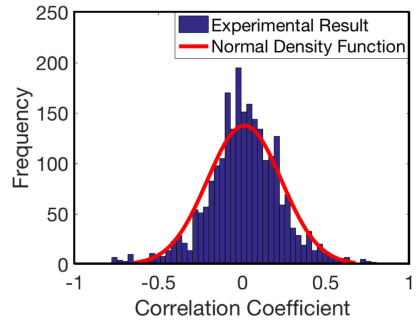
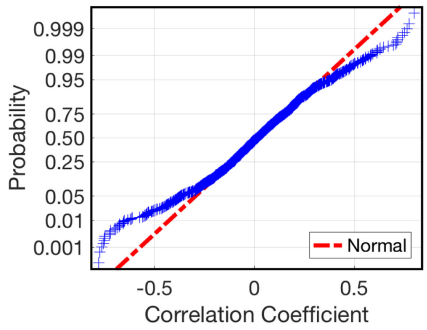


Fig. 14. The average ERP of 6 subjects. The signal in blue is evoked by the original stimuli set, and the signals in green are elicited by newly generated stimuli set. Dashed lines in red represent the duration of the stimulus hit.



(a) Histogram with a distribution fit.



(b) Normal probability plot.

Fig. 15. The correlation test between original and new ERP.

where A_i and B_j represent the original and new ERP template, respectively. In terms of the Eq. (25), normalized(A_i) consists of a_{i_n} with $1 \leq n \leq N$. Similarly, normalized(B_j) is composed of b_{j_n} with $1 \leq n \leq N$. Also, k is the total number of templates for each subject experimented on the same stimuli set, which is equivalent to 20 because each subject has 20 normalized ERP templates for one stimuli set. The correlation coefficient, R, is computed by comparing each normalized template of the old stimuli set with every normalized template of new stimuli set ($k \times k$ comparison). Then, we combine the frequency of the correlation coefficients of all subjects and graphed them into the histogram and normal probability plot.

Results and Discussions. The result of the correlation test is illustrated in Fig. 15. In Fig. 15a, our result forms the Gaussian curve centered at zero, which indicates that the mean of correlation coefficients is approximately zero. The correlation coefficient of zero signifies no linear relationship between two variables, and thus the original ERP and updated ERP are highly independent. At 95 percent confidence interval ($\alpha = 0.05$), an estimate of the mean is 0.0130 and an estimate of the STD is 0.2212. Moreover, the lower bound of the confidence intervals for the mean is 0.0040, and the upper bound is 0.0219. The lower and upper bound of the confidence intervals for the STD are 0.2151 and 0.2277, respectively. In addition to the frequency distribution histogram, Fig. 15b shows the normal probability plot to identify any substantive departure from normality. In this graph, a straight, diagonal dotted line in red provides the reference for a perfect normality. The upper end of the plot bends below the diagonal line while the lower end bends above that line, forming an S shaped-curve, which indicates a light-tailedness. In other words, our correlation results have less variance than expected. In this graph, we can also observe that approximately 90 percent of the data has a weak association because the probability from 0.05 to 0.95 ranges from the correlation greater than $-0:3$ to the correlation less than $0:3$. The strength of association is considered small for R less than 0.3 but greater than $-0:3$. Thereby, we prove the independence between the ERPs evoked by different stimuli sets and ensure that attackers are unlikely to link the old ERP to the new ERP.

9 Longitudinal Study

Objectives. Stability is essential in the biometrics-based authentication system because user's biometric signature may change over time. Thus, we conduct a longitudinal study to demonstrate the stability of ERP biometrics. In our study, both long-term and short-term performances are evaluated with the reliability change (RC) index. The long-term performance is observed to investigate whether individual's ERP signal morphs over prolonged periods of time. The short-term performance also is evaluated to analyze the effect of stimuli familiarity to the system performance.

Experiment Description. We follow the same experimental settings as Section 7.4. In the enrollment phase, we randomly select a part of owner data and use them to create a profile of the user. Then, we test the performance of the classifier by authenticating the user with the owner and all attackers. Here, we refer the authentication test in Section 7.3 as a pre-trial and re-test for a longitudinal study as a post-trial. Participants are experimented every five days expanding five months after the pre-trial. In this study, we focus on evaluating three performance metrics: *(1)* short-term accuracy (post-trial after first five days); *(2)* long-term accuracy (post-trial after about five months); *(3)* accuracy stability across the whole experimental period. For each subject, the profile of user remains the same and newly collected data are used for login attempts. Each subject acts as the owner once, and the rest acts as the attacker. This test repeats for every subject. Thus, there are total 78 tests for short-term study and 78 tests for long-term study with each test consisting 77 user attempts and 77 attacks from each attacker.

Reliability Change Index. To measure the change in the system performance over time, we calculate the reliability change index (RC) for every subject. By definition, it is a statistical method of estimating significant change. In the present study, RC index is defined by the following formula [72]:

$$RC = \frac{(X_{post} - X_{pre})}{SE_{diff}}, \tag{28}$$

$$SE_{diff} = \sqrt{2(SE_m)^2}; \quad SE_m = S_{pre}\sqrt{1 - r_{xx}}, \tag{29}$$

where pre indicates the test from Section 7.4, and $post$ represents the re-test for either short-term study or long-term study. Correspondingly, X_{pre} and X_{post} denote the individual's pre-trial and post-trial f-score. SE_{diff} is the standard error of the difference between the pre-trial and post-trial f-scores of the group, and SE_m represents the standard error of the measurement. S_{pre} is the STD of the pre-trial f-scores of the group. r_{xx} is a test-retest reliability measured by the Pearson's correlation coefficient between the pre-trial and post-trial f-scores of the group. In the current study, two-sided significance test ($\alpha = 0.05$, $z = 1.96$) is used since

TABLE 4
Overall Performance Change (f-Score)

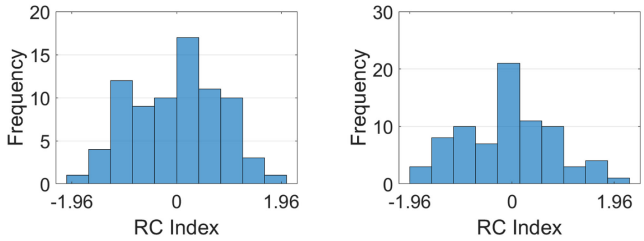
Duration	Pre-trial (%)	Post-trial (%)	Change(%)
Short-term	96.43 ± 3.99	96.45 ± 4.32	+0.02
Long-term	96.00 ± 5.81	94.99 ± 6.30	−1.01

our interest is in the performance stability. The absolute value of RC index larger than 1.96 indicates a significant performance change.

Results and discussion

(1) Long/short-term accuracy: The overall performance change is summarized in Table 4. The f-score is increased by only 0.02 percent during the short-term study, indicating that individual's acquaintance with the stimuli does not degrade our system performance. The possible reason is that our stimulus presentation is too fast to properly trigger a short-term memory, and therefore an intrinsic reaction from the semantic memory, a portion of long-term memory, overrides the response from the short-term memory. Conversely, the performance is declined by 1.01 percent during the long-term study. This change is slightly higher than the change observed in the short-term study. However, it should be noted that this change is still insignificant.

To statistically analyze the degree of change, we evaluate the RC indexes and illustrate results in Fig. 16. As observed, the frequency in y-axis indicates the number of the participants. The sum of frequencies for each graph is equivalent to the total number of subjects who have participated in the corresponding study. In Fig. 16a, all subjects for the short-term study achieves the RC index greater than -1.96 and less than 1.96. Thus, no subject experiences a statistically significant performance degradation, and this result conforms to the aforementioned analysis in Table 4. In the long-term study, only one subject has a statistically significant f-score decrease as the RC index of the subject is above 1.96. To be specific, the pre-trial f-score of this subject (X_{pre}) is 100 percent, and the post-trial f-score of this subject (X_{post}) is 90.28 percent. Yet, considering the time interval between two trials for the long-term study (142.8 days), the post-trial performance of 90.28 percent is not considerably low; it is 4.71 percent lower than the overall performance of the post-trial. Also, the rest 77 subjects have the RC index that is within the range from -1.96 and 1.96. Thus, we conclude that our system is relatively stable considering the time interval and actual post-trial performances.



(a) Short-term RC test. (b) Long-term RC test.

Fig. 16. The reliability change test results for the short-term and long-term studies. The frequency in y-axis indicates the number of the participants.

Fig. 17 visualizes the ERP signals of 6 subjects during the long-term study. We can see that a data smoothing method, similar to what we adopted for Fig. 14, is applied to depict a general trend of the brainwave without noise. The ERP signals from the pre-trial are colored in blue, and the ERP signals from the post-trial are colored in green. The dashed box in the graph represents the stimulus duration. Although there is a slight shift in position for some waveforms, the general pattern of pre-trial and post-trial ERP signals are similar. In particular, the waveforms up to 200 ms have the closest pattern than the waveforms after 200 ms. This similarity validates that an individual's brainwave does not morph rapidly and appreciably over prolonged periods of time, and shows that our ERP-based authentication system is stable.

(2) Accuracy stability: The purpose of this accuracy stability evaluation with repeated authentication is to check whether a user will get accustomed to the stimuli and develop fixed behaviors when the same stimuli were exposed frequently, which may affect the authentication performance. The accuracy stability is evaluated by measuring and comparing the post-trial f-scores every five days. The mean of f-scores measurement is depicted in Fig. 18. In the five months duration, mean values of f-scores are

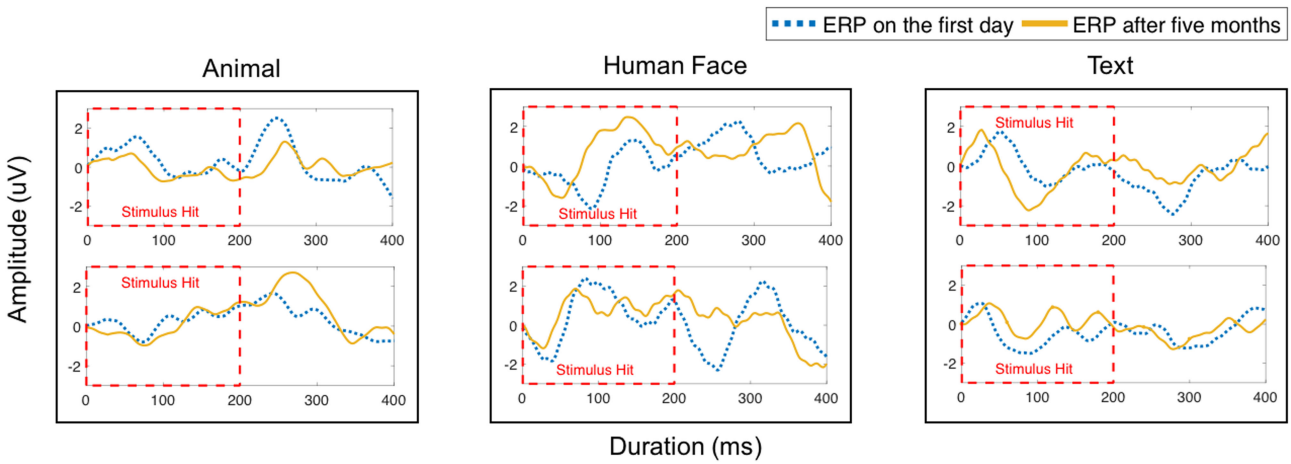


Fig. 17. The average ERP of six subjects. The signals in both blue and yellow are evoked by the same stimuli set. However, the time interval between two ERPs is approximately five months. Dashed lines in red represent the duration of the stimulus hit.

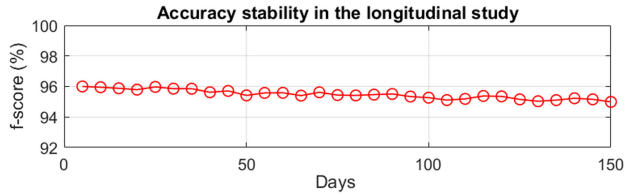


Fig. 18. The variation of the mean of f-scores across the five month longitudinal study.

between 94.99 and 96.00 percent, and the maximum variation is 1.01 percent. We can see the frequent exposure of visual stimuli and the large time gap between the first and last trials has no significant impact on f-scores during the five months experiment. Therefore, the result verifies that a user will not develop a fixed behavior towards the authentication and the ERP biometrics is stable in long-term.

10 User Experience Study

All participants are asked to fill in a questionnaire about their experience of using the ERP-based HMD system. The survey questions mainly concentrate on the comfort and the usability of the system. To ensure the objectiveness, the survey is conducted anonymously and the participants score 1 (strongly disagree) to 5 (strongly agree) for each question. Table 5 shows the statements and the average scores.

The results show an overall agreement of ERP brain signal as a secure biometric. Nearly all participants feel comfortable when wearing the headsets and are satisfied to use the system. More importantly, they really believe that the proposed cancelable biometric is more secure than the current authentication practices, and can be seamlessly integrated with VR devices. 95 percent of them are willing to see this technique commercialized as it is very secure and convenience in usage. All the high scores show the possibility of our system to be adopted by users as a new authentication framework.

TABLE 5
User Experience Questionnaire

No.	Questionnaire Statement	Score
1	*Both the ERP and VR headsets are comfortable to wear.*	4.5
2	*The time interval between images is suitable in practice.*	4.1
3	*The cancelable brain biometric is more secure than the traditional ones.*	4.5
4	*Generally, I am satisfied with the use of the ERP-based HMD system.*	4.3
5	*The ERP biometrics will become the secure and seamlessly integrated authentication for future VR platform.*	4.4
6	*I'm willing to purchase this cancelable ERP technique for more secure and convenient authentication.*	4.7

11 Discussion

Liveness Detection. To detect and prevent spoofing attacks, the authentication system must differentiate real biometrics from counterfeits. Most promising way to distinguish them is to detect physiological signs of liveness. The methods to achieve the liveness detection can be subdivided into software-based techniques and hardware-based techniques. For software-based techniques, one approach is to identify a specific characteristic of fake biometric credentials. For instance, Galbally et al. [73] distinguishes fabricated fingerprint biometrics using image quality-related measures. However, the drawback is that this kind of approach may not be applicable when original biometric signatures are simply duplicated. Another software-based approach is to request the user to provide signs of liveness or force user to interact with the system continuously. These methods, however, often decrease the user comfort. Hardware-based techniques, on the other hand, utilizes dedicated sensors to detect dynamic features in real-time. This technique is difficult to deploy with the static credentials, such as fingerprints, hand geometry, and iris unless additional hardware is added to detect other properties of living traits. Extra sensors, however, may increase the implementation cost and lower user convenience. For example, Baldisserra et al. [74] uses an odor sensor to detect liveness for fingerprint-based identification system. Although it is promising for securing fingerprint scanner, implementing such sensor adds the cost and is not feasible for smart wearables, which must be carried along with the user. In contrast, our proposed ERP-based approach is a dynamic biometric credential, which itself provides the physiological sign of liveness as the active EEG must always come from living individuals. Also, unlike odor biometrics, it is not revealable because electrodes are kept inside of the smart headwear devices and completely in contact with user's head. Additionally, since the brainwave biometrics can be obtained continuously while wearing, the system can automatically log out in the absence of brainwaves.

Aging Effect. The effects of aging on biometric authentication and the solutions to minimize these effects must be addressed. Most biometrics, such as fingerprint, iris, and face, spontaneously morph over the course of a lifetime. However, the aging process is often very slow, and the degree of alteration is negligible. Similarly, age-related alterations of brainwave have been recognized due to the overall EEG power decrease, slower alpha frequency, and slight diminution in P3 amplitude. Yet, they are negligible when ERPs are obtained from visual stimuli [75]. Thus, the signal alterations from aging should not significantly affect our system performance. Raz et al. [76] also suggest that these changes reported in many cross-sectional studies could be overestimated due to inclusion of extreme cases like person with pre-clinical dementia. Moreover, the degree of changes varies from person to person. Since such changes are particularly observed in the temporal lobe, one possible solution to alleviate these effects is to avoid the use of electrodes located in the temporal regions for authentication purpose. Moreover, as shown in the longitudinal study, a natural mutation of brain signal does not occur rapidly, therefore the ERP profile update can also be a potential solution.

Privacy Preservation. In the context of the privacy concerns, one natural question is "will this brain biometrics leak privacy information"? The answer is "No". Previous works indicate that brain leakage requires a satisfactory data, such as high-fidelity brainwaves with a professional

device (e.g., BCI2000-64 channels [9]) or invasive measures by embedding chips into brain [77]. On the contrary, our system only requires three channels with a small information disclosure. Moreover, our system only collects ERP P/N200 (i.e., within 200 ms post-stimulus onset response), while most of the semantic memory attacks require the relatively long brainwaves (e.g., non-ERP sections from seconds to minutes [78]).

12 Limitations and Future Work

In-depth Security Evaluation. Although in general, for any given visual stimuli, it is not possible for adversaries to always produce the same ERP as the victim. Whether attackers can generate similar ERPs to victim over certain visual stimuli due to the emotion synchrony (e.g., both of them just watched a same film, shared the similar emotional feelings, or experienced cooperative work process) is still unclear. In the next step, we will conduct an in-depth security evaluation on the effect of emotion synchrony towards the adversarial ERP generation.

Further Motion Artifacts Cancellation. Our ERP brainwave-based authentication system is sensitive to strong motion artifacts. To further compensate artifact noises, several motion artifact cancellation methods can be employed. While our method described in Section 4.4 is effective for non-continuous motion artifact noises, it could be vulnerable when extreme physical activities continue throughout the authentication. Thus, to counterbalance artifacts from the continuous gait events, an adaptive independent component analysis (ICA) mixture model can be applied to parse the EEG signals into maximally independent components (IC), which undergoes the component-based template regression procedure. The feasibility of this approach is proved by the data collected during the treadmill walking and running. In addition, our system can incorporate head accelerations to cancel motion artifacts. With a three-dimensional accelerometer, the system can detect artifacts induced by head movements and remove the brainwave synchronous with the recorded acceleration above certain threshold [79]. To describe in a more detailed way, head accelerations are measured relative to the initial position, and independent component analysis (ICA) identifies EEG components that are statistically independent. Then, components that correlate with the recorded acceleration above certain threshold are removed [79]. By doing so, clean signals reconstructed from the contaminated EEG data can be extracted and used for authentication credential.

Statistical and Frequency Domain Features. To further remedy the sensitivity to noise, we can include statistical features into the feature vectors. The statistical features, such as mean, median, standard deviation, root mean square, mean derivative, average first order derivative, skewness, kurtosis, interquartile range, zero crossing rate, and mean crossing rate, can be calculated at the primitive level. Since the frequency domain features are extremely correlated with variance and standard deviation, we can also employ energy, dominant frequency, and spectral entropy. Thus, 14 features need to be extracted for each channels, adding 42 features to each feature vector.

Comprehensive Evaluation. Though we have utilized the Pearson's correlation analysis for the unlinkability property assessment, we plan to provide a more comprehensive evaluation to prove that the reissued brainwave biometrics is indeed unlinkable. Specifically, Spearman's rank order correlation [80], Kendall rank correlation [81], and Hausdorff distance [82] will be employed for the analysis. At the current stage, we validated the feasibility of our brain password with 179 adult participants, a further study with a much larger sets of participants to verify the uniqueness and stability of the brain biometrics is in our plan. Another promising research direction to pursue is to investigate the impact of visual stimuli protocols, such as full color versus black and white, designated visual stimuli under other different categories.

13 Related Work

Headwear Authentication. Several authentication methods using behavioral biometrics for head-mounted displays have been researched in the past. For instance, Chauhan et al. [22] developed a touch gesture-based continuous authentication for wearable devices like Google Glass. Although such devices have limited hardware resources, their results indicate that gesture-based authentication is still feasible. Similarly, Li et al. [21] proposed an authentication system for head-worn devices using user's unique head movement patterns in response to music. Their approach was able to accurately authenticate users with an average true acceptance rate of 95.57 percent. Also, Rogers et al. [20] have presented the method to identify an HMD user based on the user's unconscious blinking and head-movement that achieves the accuracy of 94 percent. However, such physiological and behavioral characteristics can easily be observed and thus can be surreptitiously duplicated and counterfeited. Other existing techniques, such as eye movement biometrics [83], can be conveniently integrated into HMD devices. However, such physiological and behavioral characteristics are prone to compromise in daily life and thus can be surreptitiously duplicated and counterfeited.

Authentication using Brainwaves. Most past attempts at neurofeedback based authentication have used EEG as a biometric measure. For example, Chuang et al. [61] presented an authentication scheme based on single-channel EEG signals, demonstrating that brainwave could be successfully exploited for purposes of subject authentication. Similarly, Ashby et al. [60] employed EEG signals for person authentication with AR model and power spectral density. However, since regular EEG is collected without stimulation, EEG is lack of experimental control. This means that the performance of EEG biometrics is highly unstable as it depends on individual's condition at the moment of authentication. Thus, in contrast to previous attempts, we presented more efficient and feasible solution by utilizing the ERP signal, which represents a stimulus-averaged signal time-locked to a specific event. Armstrong et al. [34], Ruiz-Blondet et al. [25], and Gupta et al. [84] adopted ERP as the brain biometrics, but they never considered the cancelability of the system. While this work focuses on the biometrics cancelability including update strategy design and cancelability analysis.

Cancelable Biometric Systems. The common approaches to generate a cancelable biometric are characterized by two stages: feature extraction and transformation. Connie et al. [85] proposed a method which uses existing biometric

palmprint features with a set of pseudo-random data to generate a unique discretized code for every individual. Similarly, Paul et al. [2] developed a cancelable biometric template generation algorithm using random projection and transformation-based feature extraction for multi-modal face and ear biometrics. This approach is unique due to its use of cancelable multimodality. Further, Ouda et al. [86] exploited the feature domain transformation for protecting IrisCode. The feature transformation is accomplished by IrisCode generation, consistent bits extraction, and cancelable BioCode generation. This proposed method retains the advantages of revocability, diversity, and non-invertibility without deteriorating the recognition performance. However, these methods are based on a soft-cancellation, which generates a cancelable biometric through the alteration and transformation of existing templates. For the first time, we introduced the notion of hard-cancellation, a generation of totally new bio-features, through the manipulation of visual stimuli.

14 Conclusion

In this paper, we presented the first study to explore secure and usable authentication to headwear devices using cancelable ERP biometrics. The evaluation results show that our approach achieves the f-score accuracy of 95.72 percent, half total error rate (HTER) of 2.261 percent, and equal error rate (EER) of 2.503 percent. Thus, for the first time, we have validated the feasibility of using unique, non-volitional components of brainwave response for authentication of smart headwear users. Also, we introduced the notion of cancelability to the brainwave biometrics through a novel stimuli update strategy. A further cancelability analysis in terms of revocability and unlinkability is conducted to prove the effectiveness of the reissued biometrics credential.

Acknowledgments

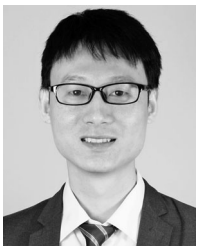
This work was in part supported by the National Science Foundation under grant No. 1266183, 1423061/1422417, 1564104/1564046, 1840790, and the National Natural Science Foundation of China under grant No. 61972348.

References

[1] C. C. C. (CCC), "Fingerprint biometrics hacked again," Dec. 2014. [Online]. Available: http://www.ccc.de/en/updates/2014/ursel, Accessed on: May 13, 2017.
[2] P. P. Paul and M. Gavrilova, "Multimodal cancelable biometrics," in *Proc. IEEE 11th Int. Conf. Cognitive Informat. Cognitive Comput.*, 2012, pp. 43–49.
[3] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 738–742, Apr. 2007.
[4] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Sci.*, vol. 150, no. 3700, pp. 1187–1188, 1965.
[5] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics*, 2011, pp. 1–7.
[6] K. Cao and A. K. Jain, "Hacking mobile phones using 2D printed fingerprints," Michigan State University, Tech. Rep. MSU-CSE-16-2, 2016.
[7] M. Poulos, M. Rangoussi, N. Alexandris, A. Evangelou, et al., "Person identification from the EEG using nonlinear signal classification," *Methods Inf. Medicine*, vol. 41, no. 1, pp. 64–75, 2002.
[8] J. Thorpe, P. C. van Oorschot, and A. Somayaji, "Pass-thoughts: Authenticating with our minds," in *Proc. Workshop New Security Paradigms*, 2005, pp. 45–56.
[9] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song, "On the feasibility of side- channel attacks with brain- computer interfaces," in *Proc. 21st USENIX Conf. Security Symp.*, 2012, pp. 34–34.
[10] D. G. Amaral, "Memory: Anatomical organization of candidate brain regions," *Comprehensive Physiology*, pp. 211–294, 2011.
[11] Samsung, "Samsung VR," 2015. [Online]. Available: http://www.samsung.com/us/explore/gear-vr/?cid=van-mb-cph-0716-10000089
[12] PlayStation, "Playstation VR," 2016. [Online]. Available: https://www.playstation.com/en-us/explore/playstation-vr/
[13] Google, "Google glass," 2016. [Online]. Available: https://www.google.com/glass/start/
[14] D. Liu, S. A. Jenkins, P. M. Sanderson, P. Fabian, and W. J. Russell, "Monitoring with head-mounted displays in general anesthesia: a clinical evaluation in the operating room," *Anesthesia Analgesia*, vol. 110, no. 4, pp. 1032–1038, 2010.
[15] J. I. Thompson, "A three dimensional helmet mounted primary flight reference for paratroopers," Air Force Inst of Tech Wright-Patterson AFB Oh, Tech. Rep., 2005.
[16] Markets and Markets, "Global head-mounted display market 2016–2020," 2015. [Online]. Available: http://www.marketsandmarkets.com/Market-Reports/head-mounted-display-hmd-market-729.html
[17] T. Rosinski, "Report on the current state of the VR market," The Farm 51, 2015. [Online]. Available: http://thefarm51.com/ripress/VR_market_report_2015_The_Farm51.pdf
[18] M. Schwartz, "Hack my google glass: Security's next big worry," *Inf. Week*, Aug. 2013. [Online]. Available: https://www.darkreading.com/vulnerabilities-and-threats/hack-my-google-glass-securitys-next-big-worry/d/d-id/1111278
[19] C. Arthur, "Google glass hacked with malicious QR code to yield its pictures and video," *The Guardian*, 2013. [Online]. Available: https://www.theguardian.com/technology/2013/jul/17/google-glass-hacked-vulnerability-image
[20] C. E. Rogers, A. W. Witt, A. D. Solomon, and K. K. Venkatasubramanian, "An approach for user identification for head-mounted displays," in *Proc. ACM Int. Symp. Wearable Comput.*, 2015, pp. 143–146.
[21] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser, "Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2016, pp. 1–9.
[22] J. Chauhan, H. J. Asghar, A. Mahanti, and M. A. Kaafar, "Gesture-based continuous authentication for wearable devices: The smart glasses use case," in *Proc. Int. Conf. Appl. Cryptography Netw. Security*, 2016, pp. 648–665.
[23] D. V. Bailey, M. Dürmuth, and C. Paar, "Typing passwords with voice recognition: How to authenticate to google glass," in *Proc. Symp. Usable Privacy Secur.*, Citeseer, 2014, pp. 1–2.
[24] K. A. Paller and A. D. Wagner, "Observing the transformation of experience into memory," *Trends Cognitive Sci.*, vol. 6, no. 2, pp. 93–102, 2002.
[25] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "CEREBRE: A novel method for very high accuracy event-related potential biometric identification," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1618–1629, Jul. 2016.
[26] E. E. Fetz, "Volitional control of neural activity: implications for brain–computer interfaces," *J. Physiology*, vol. 579, no. 3, pp. 571–579, 2007.
[27] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner, "Parietal lobe contributions to episodic memory retrieval," *Trends Cognitive Sci.*, vol. 9, no. 9, pp. 445–453, 2005.
[28] J. V. Haxby, L. G. Ungerleider, B. Horwitz, J. M. Maisog, S. I. Rapoport, and C. L. Grady, "Face encoding and recognition in the human brain," *Proc. Nat. Academy Sci. United States America*, vol. 93, no. 2, pp. 922–927, 1996.
[29] T. Klingberg, M. Hedehus, E. Temple, T. Salz, J. D. Gabrieli, M. E. Moseley, and R. A. Poldrack, "Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging," *Neuron*, vol. 25, no. 2, pp. 493–500, 2000.
[30] J. Wendt, M. Lotze, A. I. Weike, N. Hosten, and A. O. Hamm, "Brain activation and defensive response mobilization during sustained exposure to phobia-related and other affective pictures in spider phobia," *Psychophysiology*, vol. 45, no. 2, pp. 205–215, 2008.

[31] B. Jemel, M. Pisani, M. Calabria, M. Crommelinck, and R. Bruyer, "Is the n170 for faces cognitively penetrable? evidence from repetition priming of mooney faces of familiar and unfamiliar persons," *Cognitive Brain Res.*, vol. 17, no. 2, pp. 431–446, 2003.
[32] S.-K. Yeom, H.-I. Suk, and S.-W. Lee, "Person authentication from neural activity of face-specific visual self-representation," *Pattern Recognit.*, vol. 46, no. 4, pp. 1159–1169, 2013.
[33] J. W. Tanaka, T. Curran, A. L. Porterfield, and D. Collins, "Activation of preexisting and acquired face representations: the n250 event-related potential as an index of face familiarity," *J. Cognitive Neuroscience*, vol. 18, no. 9, pp. 1488–1497, 2006.
[34] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for erp biometrics," *Neurocomputing*, vol. 166, pp. 59–67, 2015.
[35] N. E. Crone, A. Sinai, and A. Korzeniewska, "High-frequency gamma oscillations and human brain mapping with electrocorticography," *Progress Brain Res.*, vol. 159, pp. 275–295, 2006.
[36] Y. N. Singh, S. K. Singh, and A. K. Ray, "Bioelectrical signals as emerging biometrics: Issues and challenges," *ISRN Signal Process.*, vol. 2012, 2012, Art. no. 712032.
[37] J. N. Mak, D. J. McFarland, T. M. Vaughan, L. M. McCane, P. Z. Tsui, D. J. Zeitlin, E. W. Sellers, and J. R. Wolpaw, "EEG correlates of p300-based brain–computer interface (BCI) performance in people with amyotrophic lateral sclerosis," *J. Neural Eng.*, vol. 9, no. 2, 2012, Art. no. 026014.
[38] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient p300-based brain–computer interface for disabled subjects," *J. Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, 2008.
[39] B. Dal Seno, M. Matteucci, and L. Mainardi, "A genetic algorithm for automatic feature extraction in p300 detection," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, 2008, pp. 3145–3152.
[40] M. E. Berryhill, L. Phuong, L. Picasso, R. Cabeza, and I. R. Olson, "Parietal lobe and episodic memory: Bilateral damage causes impaired free recall of autobiographical memory," *J. Neuroscience*, vol. 27, no. 52, pp. 14 415–14 423, 2007.
[41] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density EEG recorded during walking and running," *J. Neurophysiology*, vol. 103, no. 6, pp. 3526–3534, 2010.
[42] A. Delorme, S. Makeig, T. Jung, and T. Sejnowski, "Automatic rejection of event-related potential trials and components using independent component analysis," in *Soc. Neuroscience Abstracts*, vol. 27, pp. 457–462, 2001.
[43] S. Johansen, "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models," *Econometrica: J. Econometric Soc.*, vol. 59, no. 6, pp. 1551–1580, 1991.
[44] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN Neuroscience*, vol. 2014, 2014, Art. no. 730218.
[45] S. Jain and G. Deshpande, "Parametric modeling of brain signals," in *Proc. Technol. Life: North Carolina Symp. Biotechnology Bioinf.*, 2004, pp. 85–91.
[46] B. Friedlander and B. Porat, "The modified yule-walker method of ARMA spectral estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-20, no. 2, pp. 158–173, Mar. 1984.
[47] R. Bos, S. De Waele, and P. M. Broersen, "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 6, pp. 1289–1294, Dec. 2002.
[48] A. Alkan and M. K. Kiymik, "Comparison of ar and welch methods in epileptic seizure detection," *J. Med. Syst.*, vol. 30, no. 6, pp. 413–419, 2006.
[49] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[50] S. Hou and R. C. Qiu, "Kernel feature template matching for spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2258–2271, Jun. 2014.
[51] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.
[52] T. M. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. II–II.
[53] BetterHealth, "Eeg test," [Online]. Available: https://www.betterhealth.vic.gov.au/health/-conditionsandtreatments/eeg-test, Accessed: Sep. 17, 2017.
[54] T. M. Mitchell, *Machine Learning*, 1st ed., ser. McGraw-Hill Series in Computer Science, McGraw-Hill Education, Mar. 1997.
[55] "A technical evaluation of fingerprint scanners," Biometrika, Tech. Rep., 2015. [Online]. Available: http://www.biometrika.it/eng/wp scfing.html, Accessed: Aug. 2019.
[56] T.-Y. Jea and V. Govindaraju, "A minutia-based partial fingerprint recognition system," *Pattern Recognit.*, vol. 38, no. 10, pp. 1672–1684, 2005.
[57] S. Malathi, "An efficient approach for partial fingerprint recognition based on pores and sift features using fusion methods," Ph.D. dissertation, Avinashilingam Institute for Home Science and Higher Education for Women, Sep. 2011.
[58] "Second international competition for fingerprint verification algorithms FVC2002," [Online]. Available: http://bias.csr.unibo.it/fvc2002/databases.asp
[59] C. Watson, "Nist fingerprint data," National Institute of Standards and Technology, Gaithersburg, MD, Jan. 2015. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/Watson-Presentation.pdf
[60] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (EEG) based authentication," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, 2011, pp. 442–445.
[61] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore I am: Usability and security of authentication using brainwaves," in *Proc. Int. Conf. Financial Cryptography Data Security*, 2013, pp. 1–16.
[62] USMA, "Biometrics metrics report v.30," 2012. [Online]. Available: http://www.usma.edu/ietd/docs/BiometricsMetricsReport.pdf
[63] R. M. Bolle, N. K. Ratha, and S. Pankanti, "Error analysis of pattern recognition systems the subsets bootstrap," *Comput. Vis. Image Understanding*, vol. 93, no. 1, pp. 1–33, 2004.
[64] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[65] N. Meinshausen, "Quantile regression forests," *J. Mach. Learn. Res.*, vol. 7, pp. 983–999, 2006.
[66] D. M. Tax and R. P. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 2004.
[67] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm
[68] K. Amano, N. Goda, S. Nishida, Y. Ejima, T. Takeda, and Y. Ohtani, "Estimation of the timing of human visual perception from magnetoencephalography," *J. Neuroscience*, vol. 26, no. 15, pp. 3981–3991, 2006.
[69] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 88–100, Sep. 2015.
[70] B. Yang, D. Hartung, K. Simoens, and C. Busch, "Dynamic random projection for biometric template protection," in *Proc. 4th IEEE Int. Conf. Biometrics: Theory Appl. Syst.*, 2010, pp. 1–7.
[71] J. K. Liu and R. Steinfeld, *Proceedings Information Security and Privacy: 21st Australasian Conferenc*, vol. 9723, Berlin, Germany: Springer, 2016.
[72] N. S. Jacobson and P. Truax, "Clinical significance: A statistical approach to defining meaningful change in psychotherapy research." *J. Consulting Clinical Psychology*, vol. 59, no. 1, 1991, Art. no. 12.
[73] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Future Generation Comput. Syst.*, vol. 28, no. 1, pp. 311–321, 2012.
[74] D. Baldisserra, A. Franco, D. Maio, and D. Maltoni, "Fake fingerprint detection by odor analysis," in *Proc. Int. Conf. Biometrics*, 2006, pp. 265–272.
[75] J. Polich, "EEG and ERP assessment of normal aging," *Electroencephalography Clinical Neurophysiology/Evoked Potentials Section*, vol. 104, no. 3, pp. 244–256, 1997.
[76] N. Raz, U. Lindenberger, K. M. Rodrigue, K. M. Kennedy, D. Head, A. Williamson, C. Dahle, D. Gerstorf, and J. D. Acker, "Regional brain changes in aging healthy adults: general trends, individual differences and modifiers," *Cerebral Cortex*, vol. 15, no. 11, pp. 1676–1689, 2005.
[77] R. Q. Quiroga and S. Panzeri, "Extracting information from neuronal populations: information theory and decoding approaches," *Nature Rev. Neuroscience*, vol. 10, no. 3, 2009, Art. no. 173.

[78] R. Matovu and A. Serwadda, "Your substance abuse disorder is an open secret! gleaning sensitive personal information from templates in an eeg-based authentication system," in *Proc. IEEE 8th Int. Conf. Biometrics Theory Appl. Syst.*, 2016, pp. 1–7.
[79] I. Daly, M. Billinger, R. Scherer, and G. Müller-Putz, "On the automated removal of artifacts related to head movement from the EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 427–434, May 2013.
[80] T. D. Gautheir, "Detecting trends using spearman's rank correlation coefficient," *Environmental Forensics*, vol. 2, no. 4, pp. 359–362, 2001.
[81] H. Abdi, "The kendall rank correlation coefficient," in *Encyclopedia Meas. Statist.*, Sage, pp. 1–7, 2007.
[82] L. Wang and D. Suter, "Analyzing human movements from silhouettes using manifold learning," in *Proc. IEEE Int. Conf. Video Signal Based Surveillance*, 2006, pp. 7–7.
[83] C. Song, A. Wang, K. Ren, and W. Xu, "Eyeveri: A secure and usable approach for smartphone user authentication", in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM'16)*, San Francisco, California, Apr. 2016, pp. 1–9.
[84] C. N. Gupta, R. Palaniappan, and R. Paramesran, "Exploiting the P300 paradigm for cognitive biometrics," *Int. J. Cognitive Biometrics*, vol. 1, no. 1, pp. 26–38, 2012.
[85] T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmhashing: A novel approach for cancelable biometrics," *Inf. Processing Letters*, vol. 93, no. 1, pp. 1–5, 2005.
[86] O. Ouda, N. Tsumura, and T. Nakaguchi, "Tokenless cancelable biometrics scheme for protecting iris codes," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 882–885.



Feng Lin (S'11-M'15) received the PhD degree from the Department of Electrical and Computer Engineering, Tennessee Technological University, , in 2015. He is currently a professor with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University, China. He was an assistant professor with the University of Colorado Denver, a research scientist with the State University of New York (SUNY) at Buffalo, and an engineer with Alcatel-Lucent (currently, Nokia). His current research interests include mobile sensing, Internet of Things security, biometrics, AI security, and IoT applications. He was a recipient of the Best Paper Award from the 2017 IEEE BHI Conference, the Best Demo Award from the 2018 ACM HotMobile Conference, and the First Prize Design Award from the 2016 International 3D printing competition. He is a member of the IEEE.

Kun Woo Cho (S'18) received the BS degree in computer engineering from the State University of New York (SUNY) at Buffalo. She is currently working toward the PhD degree in the Department of Computer Science, Princeton University, under the direction of Prof. K. Jamieson. Her research interest focuses on wireless systems, IoT, and deep learning. She is a student member of the IEEE.

Chen Song (S'14) received the BS degree in optic science and engineering from Fudan University, and the MS degree in electrical engineering from the State University of New York at Buffalo, where he is currently working toward the PhD degree with the Department of Computer Science, under the direction of Prof. W. Xu. His current research focuses on smart manufacturing, mobile health, and emerging biometrics. He is a student member of the IEEE.

Zhanpeng Jin (S'07-M'10-SM'15) received the BS and MS degrees in computer science and engineering from Northwestern Polytechnical University, and the PhD degree in electrical engineering from the University of Pittsburgh, in 2010. He was a postdoctoral research associate with the University of Illinois at Urbana-Champaign (UIUC) and an associate professor at Binghamton University, State University of New York (SUNY). He is currently an associate professor in computer science and engineering with the University at Buffalo, SUNY. His research interests include emerging biometrics, mobile and wearable computing, IoTs and cyber-physical systems, ubiquitous sensing, and neuromorphic computing. He has served as the associate editor for the *Computers and Electrical Engineering*, *Computers in Biology and Medicine*, and *BioMedical Engineering Online*. He is a senior member of the IEEE.

Wenyao Xu (M'13-SM'19) received the bachelor's and master's degrees from Zhejiang University, China, and the PhD degree from the University of California at Los Angeles, Los Angeles. He is an associate professor with tenure in the Computer Science and Engineering Department, University at Buffalo (SUNY). His research has focused on exploring novel sensing and computing technologies to build up innovative Internet-of-Things (IoT) systems for high-impact human-technology applications in the fields of smart health and cyber-security. Results have been published in peer-reviewed top research venues across multiple disciplines. To date, his group has published more than peer-reviewed 160 papers, won seven best paper awards, two best paper nominations, and three international best design awards. His inventions have been filed within the U.S. and internationally as patents, and have been licensed to industrial players. His research has been reported in high-impact media outlets. Currently, he serves as an associate editor of the *IEEE Transactions on Biomedical Circuits and Systems* (*TBCAS*), the technical program committee of numerous conferences in the field of smart health and Internet of Things, and has been a technical program committee co-chair of IEEE Body Sensor Networks in 2018. He is a senior member of the IEEE.