Two-Factor Authentication System Using P300 Response to a Sequence of Human Photographs

Netiwit Kaongoen, Moonwon Yu, and Sungho Jo, *Member, IEEE*

***Abstract*—This paper proposes a two-factor authentication system that utilizes the knowledge factor: the knowledge of client's acquaintances as the key and inherence factor: P300 ERP responses to the visual stimuli as the medium. The system works by presenting a sequence of human photographs consisting of random people photographs mixed with a few of client's acquaintances photographs that trigger P300 responses. The system then verifies the client by considering the correctness of P300 responses to the client's acquaintances photographs. The proposed system achieves an error rate of nearly zero outperforming other brain-signal-based systems and has advantages over other conventional systems in the situations where the key is exposed to the imposer.**

***Index Terms*—Authentication, biometric, computer security, electroencephalography, P300.**

I. Introduction

PEOPLE need security systems to protect their possessions and personal information. Today, many services have been developed to provide security for customers' assets, authorities, and private information. These security services are equipped with authentication systems to prove the ownership of clients and ensure that their properties are accessed rightfully.

Authentication is a verification process to decide whether to accept or decline the claimed identity of a client. There are three types of authentications based on three different factors: 1) knowledge factor (something user knows); 2) ownership factor (something user has); and 3) inherence factor (something user is) [1]. A knowledge-based authentication system includes user-input password, challenge response, and security question. An ID card, key, credit card, and security token are examples of ownership-based authentication systems. Finally, the methods requiring the identification of fingerprints [2], palm-print [3], iris [4], face [5], and other biometric identifiers are inherence-based authentication systems.

However, the current authentication systems are susceptible to being deceived by attackers. For example, knowledge-based authentication can simply be broken through the shoulder surfing method. Stealing or forging an ID card or a key is a very direct and effective method to break an ownership-based system. For inherence-based systems, methods that forge the client's identity or trick the authentication system such as using a gummy finger to forge the client's fingerprints [6] or fake biometric detection [7], [8] that uses the image processing techniques to fool the system can be used to break the system. In this paper, we will term this type of attack in which the attacker owns the client's "key" that is necessary to authorize the system (whether it is an item such as the actual key, knowledge such as password, or fake inherence factor, such as the gummy finger) the "unveiled attacks." This type of attack is the most problematic one because most of the current authentication methods are helpless against the unveiled attack and once the key is unveiled to the attacker, the success of attack is almost guaranteed. Therefore, an authentication system which can effectively withstand the unveiled attack is indispensable to properly protect the client's private properties.

Using the brain signal for the inherence-based authentication system is a great alternative because it could reduce the risk of getting the key forged since it is impossible to steal someone's brain signal and copying or making fake brain signals is unavailable using current technology. In addition, a brain is least likely to be damaged than other organs. Therefore, researchers have recently suggested the use of personal brain signals [9]–[13] to satisfy the need of new authentication systems that can protect the system against the unveiled attack.

However, brain signal-based authentication systems are not yet applicable in reality. First, brain signals can easily change unlike other inherence factors. They vary with the mental and physical state of a person, causing the instability of the system. Moreover, collecting one's brain signals, as done in the previous studies, is a time-consuming job; it requires the user to wait for more than a minute to collect the brain signal data. Further, the brain signal used in the system are acquired when the user performs some form of mental activities, for example, imagining some movements, rotating some object,

Manuscript received March 10, 2017; revised April 28, 2017, July 26, 2017, and September 15, 2017; accepted September 17, 2017. This work was supported in part by the ICT Research and Development Program of MSIP/IITP (Development of non-invasive integrated BCI SW platform to control home appliances and external devices by user's thought via AR/VR interface) under Grant 2017-0-00432, and in part by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT under Grant NRF-2016M3C7A1914448. This paper was recommended by Associate Editor P. Chen. *(Netiwit Kaongoen and Moonwon Yu contributed equally to this work.) (Corresponding author: Sungho Jo.)*

The authors are with the School of Computing, KAIST, Daejeon 34141, South Korea (e-mail: netiwit.k@cs.kaist.ac.kr; moonwon.yu@cs.kaist.ac.kr; shjo@kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2017.2756673

or thinking about something which, for some users, might be difficult and require some time to practice for the system to work. This is burdensome for the user to use the system.

The objective of this paper is to solve the problems of the previous brain-signal-based authentication systems and to provide users with a practical system, which can be used in real life. We present a two-factor authentication system by combining the merits of the inherence and knowledge factors. Among the methods for measuring brain signal, we chose the P300 event-related potential (ERP). P300 ERP elicits when the subject observes the target stimuli from the sequence of stimuli from two classes: 1) target and 2) nontarget. By using this property of P300 ERP, we can set the difference between the two classes as the knowledge factor and construct a two-factor authentication system. In this paper, pictures are used as the visual stimuli for the P300 ERP and the knowledge factor: client's personally registered pictures which can be distinguished from the nontarget pictures only by the client are employed as the key. Therefore, in order for a person to authorize himself using the proposed system, he must, first, be able to distinguish the target pictures from the nontarget pictures which will make the P300 ERP to elicit at the correct time and, second, his P300 ERP must match with the clients P300 ERP that are registered to the system. To demonstrate the feasibility of our system, we conducted experiments in three scenarios including the one dealing with the unveiled attack, which is the most vulnerable circumstance in an authentication system. The detailed methods of the proposed system and the experiments will be described in the following sections.

II. Background

A. Brain-Signal-Based Authentication System

Authentication systems that utilize signals from the brain have been suggested in the previous studies. Although each previous system varies in the details of the methods, all systems generally acquire the electroencephalogram (EEG) samples from the client, register them to the database, and use the machine learning techniques to compare the EEG from a person who wants to authorize himself with those EEG samples of the identity stored in the database. Since the EEG is not stable like other inherence factors, brain-signal-based authentication systems usually use the EEG acquired when the user are performing some kind of task. There are two methods shown in the previous studies. The first method which is used by most of the previous studies is to acquire the EEG while the user performs a mental activity solving mathematical problems, composing a letter, or imagining body movements [9]–[11]. Although the systems that use EEG signals from mental activities commonly give high accuracy, there are disadvantages that make this type of system not practical for real application. The mental activity that the users are required to perform might be a burden to the users. These systems usually have long performing time because of the mental activity and they also require some practice from the users before the acquired EEG is as intended.

The second method is to use the EEG that reacts to the visual or auditory stimuli that the system provides. For instance, the authentication system in [12] utilizes the visually evoked potentials (VEPs) together with the electrooculogram from eye-blinking acquired when the user is presented with visual stimuli. The systems that use this "reactive" method are more user-friendly and require less effort from the user since the user can just sit, observe the stimuli, and let the system do all the work. However, the drawback of the reactive method is that, although different people have different brain signal reaction to the given stimuli, the characteristics of the reaction are still the same. Take steady state VEP (SSVEP) as an example. When each subject observes the visual stimuli that flicker in a specific frequency, all of the subjects show the increment in power of the EEG signal in the frequency domain at that specific frequency, even though each individual subject has a different rate of increment in power [14]. In some cases, the differences between the reactions of EEG between each individual are not big enough and that might cause the system to misclassify the EEG data. With this reason, the reactive method is not suitable for the authentication system that requires high accuracy and thus, there are only a few studies that proposed the system with the reactive method.

Keeping the advantages and disadvantages of the reactive method described above in mind, we decide to choose the P300 ERP as the inherence factor in our system. Not only that the P300 ERP is relatively consistent within an individual and requires shorter measuring time compared to the other types of brain signal from the previous brain signal-based authentication systems [14], the characteristic of how P300 ERP is elicited allows us to construct a two-factor authentication system that use the knowledge factor as the second layer of protection which can help solving the problem of brain-signal-based authentication system that use the reactive type of brain signal. In fact, the system in [13] conducts a study that attempts to identify the subjects based only on the difference in the characteristics of P300 response between different people and the result shows a high rate of misclassification. The experiments conducted in this paper will show that by constructing the two factors together in a single system, we can achieve a system with low error rate while maintaining the advantages of the reactive method which is more convenient for the user to use and thus, suitable for real-life application.

In addition, this is an extension of our previous work [15] which presented a preliminary study using P300 ERPs responded to the photographs of user's acquaintances. The improvements of the current study compared with the work in [15] includes: 1) the number of subjects that performed the experiments increases from four subjects to ten subjects; 2) the number of trials in each experiment scenario increases from 10 to 15 trials; 3) the EEG acquisition tool is changed from wet electrodes to dry electrodes with custom-made 3-D-printed headgear to make the subject more comfortable when using the system; 4) the classification method have changed from method that uses averaged classifier outputs to accept the authorization to ranking and threshold method (see Section III-G for detailed descriptions); 5) instead of authorizing a client using the P300s from 1 and 5 blocks of stimuli only, this paper has increased the number of block from 5 to 30 and authorize a client using all number of blocks from

1 to 30 to see the trend of how the error rate changes when we increase the number of stimuli; and 6) the results analysis have been done in a more comprehensive way, for example, this paper also tries taking out the knowledge factor so that only P300 ERPs are used to calculate the results to compare and prove that the proposed two-factor system have superior performance and can solve the problem of the previous studies.

B. P300

A P300 has been widely used in various types of applications in the brain computer interface (BCI) community. A P300 is a positive ERP in EEG that occurs approximately 300 ms after the onset of target stimuli that are presented according to the Oddball paradigm which is an experimental design in which the sequence of stimuli is composed with high-probability nontarget stimuli and low-probability target stimuli. P300 is usually found together with N200 component which is the negative ERP occurring approximately 200 ms post-stimulus. The amplitude of P300 ERP may depend on the type of stimulus and the person. The latency of P300 also varies from 200 to 750 ms between different individuals [17], [18].

C. Acquaintances Photographs As Private Key

Among different types of picture, human photographs are chosen as the visual stimuli in this paper. Every person is a part of multiple social groups such as family, religion group, hobby group, school, college, and workplace. Although some of a client's acquaintances might be known to the others, especially when clients are close to each other or from the same social group, it is unlikely that one would know all the client's selected acquaintances when they are properly chosen from different social groups.

It has been shown in many studies that the human face is a reliable visual stimulus for P300-based BCI paradigm. The study in [19] developed an application for smartphone that detects the P300 ERP when the phone flashes the photograph of the person whom the user wishes to dial. The study in [20] has shown that the changes in facial pattern and facial expression in a dummy face can be used to elicit the P300 ERP which results in a high accuracy and information transfer rate in the offline single trial classification. In addition, the study in [21] and [22] also shows that using the occurrences of human faces as the visual stimuli in P300-speller matrix can significantly increases the classification accuracy and information transfer rate comparing to the conventional P300-speller method that use characters flashing as the visual stimuli.

A subset of photographs of clients' acquaintances is used as the key in our authentication system. The clients are asked to select five acquaintances from distinctively different social groups and register their photographs to the system. All photographs used in the system are taken with the same orientation (front view with face straight up to the camera) and set to have the same size and position. From this knowledge factor, the system decides to accept or reject the claimed identity depending on whether the client is able to recognize the acquaintances of the claimed identity.

III. Method and Experiment

Our authentication system uses two identity factors: 1) knowledge and 2) inherence. The details are described as follows.

A. Proposed Authentication System

Our authentication system uses the P300 ERP and photographs of acquaintances to prove the client's identity. Human photographs are used as stimuli for the P300 ERP; photographs of the client's acquaintances were used as the target stimuli, whereas photographs of random persons were selected as nontarget stimuli. In the proposed system, a small number of target stimuli were mixed with a large number of nontarget stimuli according to the Oddball paradigm to allow the target stimuli to trigger the P300 response in the client's brain. The remarkable point of this system is that the Oddball paradigm is a secret that only the real client knows. In other words, only the true owner of the claimed identity would show the P300 response when observing the target stimuli; thus, the system can detect these P300 responses and authorize the client.

Fig. 1 shows the entire scheme for the proposed authentication system. The user who wants to use the authentication system has to register to the system by providing a user identity and N target photographs [Fig. 1(a)]. The system then generates sequences of photographs consisting of random persons and target people according to the Oddball paradigm. The EEG signals from both target and nontarget stimuli are processed and used to train the P300 classification model through a supervised machine learning technique. Finally, the system stores user's target photographs and the trained P300 classification model for user's data and labels it with the user's identity.

Fig. 1(b) shows the schematic of a client using the proposed authentication system. For an identity-unknown client, Y, with claimed identity ID_X, the system loads the database of the identity ID_X and generates a random sequence of photographs using the target photographs of X that are stored in the database. Next, the system acquires EEG signals of Y and verifies the P300 responses with the P300 classification model of X. If Y is the real client, P300 response occurs in reaction to the target photographs of X, and the authentication system accepts the claim of Y.

B. Experimental Composition

We conducted the experiment in four parts: 1) training; 2) self-authentication; 3) veiled attack; and 4) unveiled attack. The P300 classification model is trained using the data collected from the training part of the experiment.

Consider the following scenario for better understanding of the other three experiment parts. There are three people: Alice, Bob, and Eve. They all are clients of a bank and have their personal possessions secured in the same bank. One day, Alice goes to the bank and asks a teller for access to her safe (self-authentication). Meanwhile, Bob desires for Alice's possession, and thus asks a teller for access to Alice's safe (veiled attack). Eve is more malicious than Bob. She has secretly investigated Alice and succeeded in collecting all information

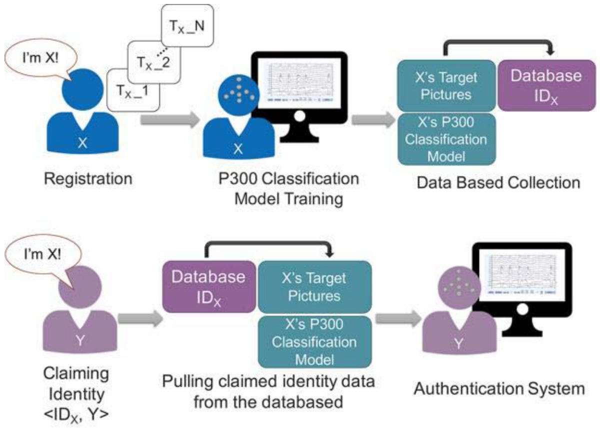

I'm X!
T_X_N
T_X_2
T_X_1
X
Registration
X
P300 Classification Model Training
X's Target Pictures
Database ID_X
X's P300 Classification Model
Data Based Collection
I'm X!
Y
Claiming Identity <ID_X, Y>
Database ID_X
X's Target Pictures
X's P300 Classification Model
Pulling claimed identity data from the databased
Y
Authentication System

Fig. 1. Schemes for the authentication system. Top: registration scheme. Bottom: utilization scheme.

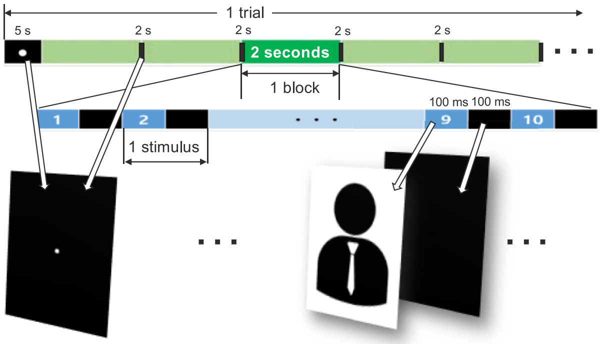


Fig. 2. Temporal scheme of the experiment. Each trial of the experiment starts with 5 s of black screen. Each block in a trial is separated with 2 s of black screen. The visual stimuli are presented for 100 ms and the ISI is of 100 ms.

needed for authentication. Eve memorizes Alice's information, goes to the bank and asks a teller for access to Alice's safe (unveiled attack).

Thus, from the attacker's viewpoint, there are two types of attacks. We define the following terminologies to categorize the types of attacks in the authentication system.

1) *Veiled Attack:* The attack method in which attackers know the target's basic information, such as age, gender, characteristic but the user's important information to pass the security such as a key is veiled.
2) *Unveiled Attack:* The attack method in which attackers know all necessary information to break the authentication system.

C. Subjects

Ten healthy test subjects (aged 26.9 ± 2.0; consisting of three females and seven males) participated in our experiment. They all gave written consent. The KAIST Institutional Review Board approved the proposed experimental protocol of this paper. All the test subjects were free of any neurological disorder and eye problem, and had never experienced any type of brain signal-based authentication system experiment. Further, they had no problem in understanding the procedure of the authentication system and knew what they needed to do during the experimental tasks. The experiments were conducted in a quiet and comfortable environment. The subjects were asked to wear earplugs and the EEG acquisition device and sit in front of the PC monitor during the experiment.

D. Experimental Setup

Fig. 2 depicts the entire experimental process in detail. Each trial starts with a standby screen (black screen with a small white fixation point in the center), which lasts for 5 s. Each trial is composed of 30 blocks of stimuli, which contain 10 randomly ordered photographs in which only two were target photographs randomly drawn from client's database and the other eight are photographs of random persons. Each photograph stimulus is shown in the screen for 100 ms, followed by a black screen for another 100 ms [i.e., the interstimulus interval (ISI) is 100 ms]. Two consecutive blocks are separated with 2 s of standby screen. Every photograph including the standby screen is 300 × 400 in size.

In the veiled and unveiled attacks of the experiment, the test subjects are asked to deceive the system by being an attacker to another subject's system; they pretend that they are the owner of the claimed identity. For each subject, one of the other subjects is chosen as the objective of the attack (the objective subject of the attack is different for all subject). The veiled attack starts first by letting the subject try to authorize himself to the objective system without any information about the claimed identity and then the unveiled attack task starts by having the test subject memorize all the target pictures of the claimed identity and try to authorize himself again to the objective system.

E. Data Acquisition

EEG data were recorded in eight channels, including Fz, Cz, Pz, P3, P4, Oz, PO7, and PO8 according to the International 10–20 system by using OpenBCI [23], 32 bit board kit with a sampling rate of 250 Hz. All the channels are referenced and grounded to the left earlobe (A1) using an ear-clip. We 3-D-printed a headgear by using the adaptation from the Ultracortex Mark 3 model provided by OpenBCI (Fig. 3). All the electrodes for each EEG channel are Ag–AgCl dry spiky electrodes [24]. The custom-made 3-D-printed headgear can be adjusted in size to fit each subject perfectly. The dry electrode and customized headgear allow the subjects to perform the experiment conveniently.

F. Data Processing

EEG data from each block are segmented into EEG epochs, with lengths of 800 ms, from the onset of each stimulus to represent the stimulus. Note that in each EEG epoch, the last 600 ms overlap with the data of the next EEG epoch from the subsequent stimuli.

Every EEG epoch undergoes the following processing methods to improve signal-to-noise ratio and remove the artifacts before further analysis. First, EEG data are filtered using a fourth-order Butterworth bandpass filter. The cutoff frequency is set to 1.0–12.0 Hz. Second, EEG data in each epoch are decimated with a factor of 12 by using Chebyshev

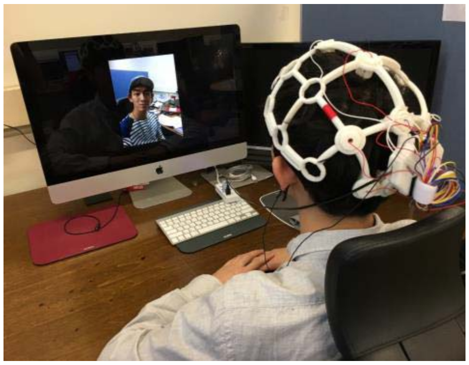


Fig. 3. Subject wearing custom-made 3-D-printed headgear with OpenBCI 32-bit board kit while performing the experiment.

Type I lowpass filter to reduce the size of feature vectors. Winsorization is then performed by computing the tenth and ninetieth percentiles for each EEG channel and the values lower or higher than these percentiles are replaced by the tenth and ninetieth percentile values. Next, EEG data from each channel are normalized to have zero mean and standard deviations equal to one. Finally, the feature vectors for each EEG data epoch are constructed by concatenating samples from each of the eight channels into one data vector. Furthermore, the vector is constructed to label each EEG data epoch with $(-1, 1)$ to indicate whether the EEG epoch is the response for nontarget or target stimuli (i.e., non-P300 or P300 class).

G. Classification Method

We used Fisher's linear discriminant analysis (FLDA) to construct the P300 classification models. FLDA computes a discriminant vector that separate the two classes: 1) P300 response and 2) non-P300 response such that the distance between the projected means of two classes is maximized while the within-class variance is minimized [25]. For the binary classification problem as used in this system, the discriminant vector w can be computed as

$$w = \left(X^T X\right)^{-1} X^T y \tag{1}$$

where X is the matrix of features vectors from training data and y is the vector of class labels $(-1, 1)$. The output of FLDA given an input vector $\hat{x}$ is simply the product $w^T \hat{x}$. FLDA have been proved that it provides better overall performance for practical P300 classification comparing to other common classifiers [26]. To fix the imbalance in the number of samples between P300 and non-P300 classes, the system uses a random over-sampling method to increase the number of P300 sample to as same as the number of non-P300 sample. This method has proved that it is effective in P300 classification [27]. The authentication system stores clients' classification model trained by their respective training data. The trained model detects a client's P300 ERP from brain waves by estimating the probability of the brain wave being the P300 ERP response to the target stimuli. A feature vector constructed from the EEG epochs responded to each visual stimulus is used as the input of the classification model. Every block consists of ten EEG epochs (from two target and eight nontarget stimuli). After classification, ten EEG epochs are ranked according to the classification outputs (probability of the input EEG data being the P300 response).

Let us define each attempt to authenticate the claiming identity as a trial. The score of accepting the claimed identity in each trial is represented by $\text{RANK}_{\text{trial}}$, which is the average of the ranks obtained from target-labeled EEG epochs. $\text{RANK}_{\text{trial}}$ is computed as follows:

$$\text{RANK}_{\text{trial}} = \frac{\sum_{i \in T_{\text{trial}}} \text{rank}_i}{|T_{\text{trial}}|} \tag{2}$$

where T_{trial} is a set of target stimuli in a trial (i.e., $|T_{\text{trial}}| = 2n$, where n is the number of block used to calculate the result in a trial) and rank_i denotes the rank of an EEG epoch from target stimulus i. Finally, the system decides to accept the client if and only if $\text{RANK}_{\text{trial}} = \theta$, where θ is the rank threshold. Equation (3) represents the final decision of the proposed authentication system using indicator function I

$$\text{Accept(trial)} = I(\text{RANK}_{\text{trial}} = \theta). \tag{3}$$

H. Evaluation

The proposed authentication system was evaluated with two error rates: 1) false rejection rate (FRR) and 2) false acceptation rate (FAR). Both error rates are calculated using the following formula:

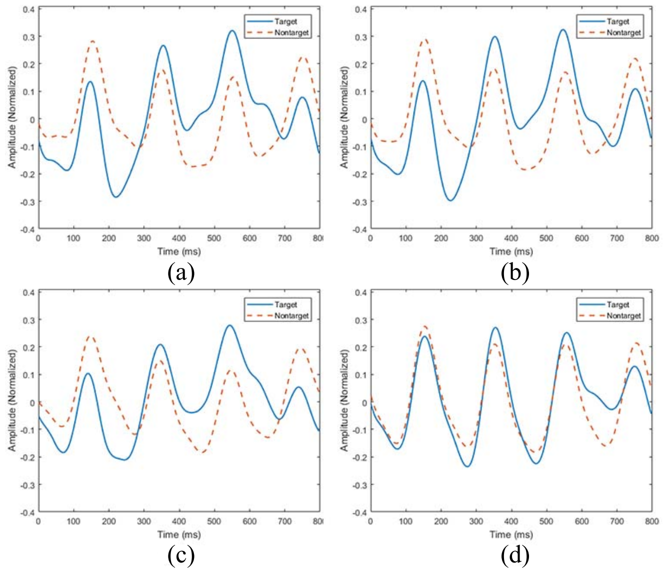
$$\text{FRR} = \frac{\text{\#Client incorrectly detected as Imposer}}{\text{\#trial}} \tag{4}$$

$$\text{FAR} = \frac{\text{\#Imposer incorrectly detected as Client}}{\text{\#trial}}. \tag{5}$$

FRR was calculated using the data obtained from self-authentication and FAR was calculated using data obtained from both veiled and unveiled attacks in the experiment.

IV. Results and Discussion

The preprocessed EEG epochs of each subject were averaged and plotted to compare between the response to the target and nontarget stimuli. The averaged EEG epochs of each subject show P300 component with similar latency in all tasks of the experiment although the amplitude are slightly varied between the tasks (especially the unveiled attack task) for some subjects. Fig. 4 shows the grand averaged EEG epoch across all subjects from each task of the experiment. The solid line indicates the EEG signal of the subjects after being presented with target stimuli and the dotted line indicates the EEG signal after being presented with nontarget stimuli. From the figure, we can clearly see the higher amplitude of the EEG from approximately 300 ms to 700 ms post-stimulus and the lower amplitude at approximately 200 ms post-stimulus in the target condition indicating the occurrences of P300 and N200 ERPs in the EEG obtained from the registration session, self-authorization, and unveiled attack task of the experiment although the amplitude of ERPs in the unveiled attack task is slightly lower compared to the other two tasks. In the veiled attack part, P300 and N200 ERPs are not observed in the EEG epochs from both target and nontarget stimuli. In addition,



(a) (b)

(c) (d)

Fig. 4. Comparison between the grand averaged filtered EEG epochs obtained when the subjects are presented with the target stimuli (solid) and non-target stimuli (dotted) from (a) registration session, (b) self-authorization, (c) unveiled attack, and (d) veiled attack task.

analysis of variance (ANOVA) is also performed to statistically confirm the difference between the EEG epochs when the subject observes the target and nontarget stimuli and the occurrences of the P300 ERP. We used ANOVA to compare the maximum amplitude in the EEG epoch by setting the null hypothesis that the mean of the maximum amplitude is statistically equal for the two populations (EEG epochs when the subject observes target stimuli and EEG epochs when the subject observes nontarget stimuli). The null hypothesis will be rejected if the P300 ERPs occur and cause the EEG epochs in target condition to have higher maximum amplitude. The results show that the null hypothesis for data from all parts of the experiment except the veiled attack part are rejected with p-value < 0.05 while the null hypothesis for the data from veiled attack part cannot be rejected and it shows high p-value (p-value > 0.5). The data from all subjects show the same results. These results are similar for all of the subjects.

Fig. 5 summarizes the performance of the authentication system. The self-authenticating part was evaluated using FRR, and the attack parts were evaluated using FAR. Our system accepts the claimed identity when the averaged rank across the target-labeled EEG epochs does not exceed the threshold rank θ. As shown in Fig. 5, the FRR decreases and FARs increase with respect to the value of θ. The high variance of FRR in self-authenticating part indicates the instability of the system across the subjects, however, the value of variance decreases as we increase the value of θ and becomes very low when the threshold is above 4. The graphs show that FRR and FAR plots have a tradeoff relationship and tend to cross when θ ranges from 4.1 to 4.5 regardless of *n:* the number of blocks. From the security viewpoint, FAR is more important than FRR because FAR indicates that the authentication system accepts false identities. The rank threshold θ could be selected depending on this oversight of security. For example, one reasonable choice for the threshold θ is 4.3 because it is the value for which both error rates from all three parts of the experiment cross and the value of all error rates and also their variance are almost zero.

Fig. 6 shows the error rates of the system depending on n. The error rates decline rapidly as we increase the value of n. The system was able to achieve an FRR of 0.000 with FAR of the veiled and unveiled attacks equal to 0.003 and 0.010, respectively, when $n = 20$. Although the accuracy of the system increases with more blocks of stimuli, it would take longer for the system to authenticate a client. One block of stimuli has a length of 4 s including 2 s of resting time, separating each block. Thus, one should consider this tradeoff when using the proposed system in real-life applications.

When $\theta = 4.3$ and $n = 15$ (1 min to authenticate a client), FRR $= 0.010$ and FARs of veiled and unveiled attacks are 0.007 and 0.011, respectively. In the scenario mentioned in Section IV, it can be said that Alice will pass our authentication system with 99.0% accuracy (FRR $= 0.010$). However, it is almost impossible that Bob is authenticated as Alice by the system (FAR $= 0.007$). Moreover, although Eve knows all the information about Alice, it is still difficult to breach the security (FAR $= 0.016$).

Although the FAR of the unveiled attack experiment is slightly higher than that of the veiled attack experiment, this outcome is still sufficient to protect the personal possessions from extreme conditions in which the key is exposed to the attacker. In other conventional single-factor authentication systems such as the password-based system, an unveiled attack would result in FAR of 1.00 because the key is the only thing you need to authenticate yourself. By using brain signals as the medium for authentication, the system is prevented from sequential attacks (e.g., brute force attack). Without brain signals as the second factor of the authentication system, an attacker would have a 2% chance [$1/\binom{10}{2}$, chance to select two target photographs from 10 photographs] of launching a brute force attack on our system, thus showing that the use of brain signals can alleviate the weakness of knowledge-based authentication.

To illustrate that the proposed method has better performance than the single-factor authentication using only P300 as the inherence factor, we use the P300 responses of one subject as inputs to the P300 classification model of another subject and calculate the FAR. We call this setting the P300-only method, and it is equivalent to the ideal case of unveiled attack in which the attacker has perfect responses to the target stimuli. We then plot the FAR results from P300-only method to compare with FARs from the veiled attack experiment in our proposed system. The result in Fig. 7 shows that our two-factor method significantly outperforms the P300-only method with respect to both accuracy and variance. The high value and variance of FAR in P300-only method indicates that different subjects might have similar pattern in P300 response causing the P300 classification model of a client to recognize other clients as the owner. This proves that the authentication

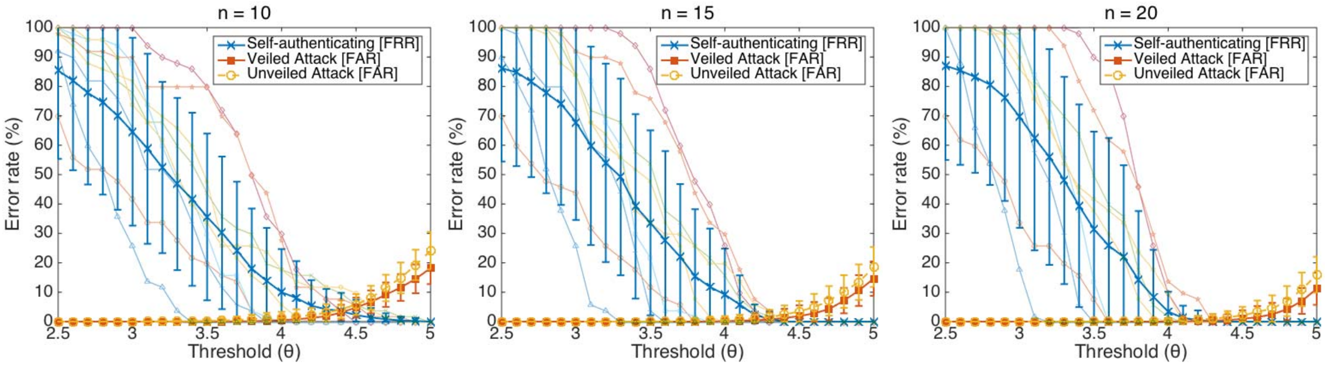


Fig. 5. Performances of the proposed system. Cross-mark line indicates FRR of self-authentication; square- and circle-mark lines indicate FAR of veiled and unveiled attack parts, respectively. From left to right, the numbers of blocks used in the trial are 10, 15, and 20 blocks, respectively. The three lines in all the three graphs are crossed at approximately $\theta = 4.3$, where all error rates are close to zero.

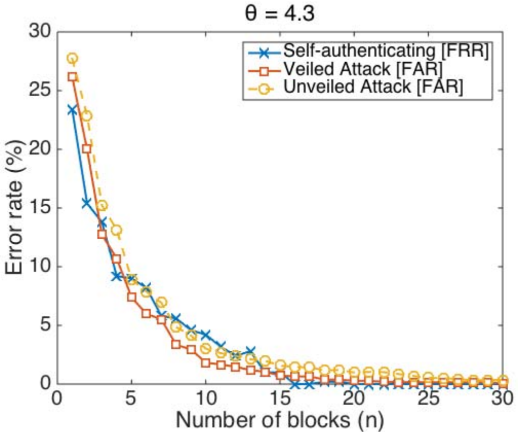


Fig. 6. Error rate (%) results from all three settings of the experiment. The error rate in all settings decreases when using more number of blocks in a trial.

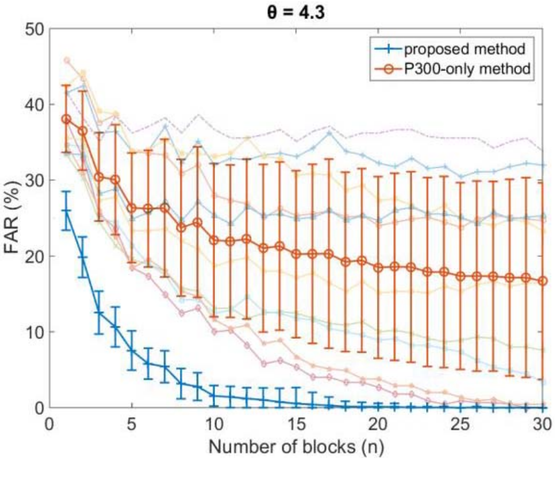


Fig. 7. Comparison between FARs of veiled attack in our proposed method and EEG only method.

system is more secured when we combine knowledge and inherence factors into a single two-factor authentication system.

In addition, the custom-made 3-D-printed headgear was easily wearable and removable unlike the wet electrode type that was used in our previous preliminary study. The size of the headgear and location of the electrodes are adjustable to make it fit to each individual. No subject reported any discomfort or any difficulty in using the proposed authentication system.

One of the concerns for the proposed system is that it might not be applicable for users with face-blindness (inability to distinguish the face) who might not be able to differentiate the target and nontarget stimuli hence, no or low P300 response. However, the proposed system is adaptive with the choice of visual stimuli. Human photographs can be changed to other categories of pictures, for example, the target stimuli might be the scenery pictures that the client has taken by himself and the system will works the same as long as the target stimuli are distinguishable only by the client. Another concern is that the client might know some of the random pictures causing a false P300 response. This problem can be solved by increasing the amount of random picture in the database or the picture that are used for the nontarget stimuli can be artificially made. The problem of BCI illiteracy that some people might not be able to use some specific type of BCI [28] which is P300 in this case is also one of the concerns. This problem can be solved by incorporating the current system with another type of BCI such as SSVEP (possibly by adding a mark that flickers in different frequency in each visual stimulus) into a hybrid system.

In conclusion, the experiment conducted in this paper have shown that our two-factor authentication system have similar or better performances than the previous brain-signal-based authentication systems [9]–[12], [15] except the system presented in [13] which shows perfect accuracy in FRR and FAR from most of the subjects. However, our system requires less effort from the user in both training and performing sessions and also time-convenient to the user. This concludes that by combining the knowledge factor with the P300 ERP as the inherence factor, the system can achieve an efficient performance while still maintaining the advantages of the brain-signal-based authentication system that use the reactive method to acquire the brain signal. We believe that the approach we have proposed here is superior to other previous single-factor brain-signal-based authentication systems and with minor adjustments it will be suitable and practical for real-life use.

References

[1] Federal Financial Institutions Examination Council. (Aug. 2001). *Authentication in an Internet Banking Environment. Financial Institution Letter, 14*. [Online]. Available: www.ffiec.gov/pdf/authentication_guidance.pdf

[2] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition for uncontrolled pose and illumination changes," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 1, pp. 149–163, Jan. 2013.

[3] B. Zhang, W. Li, P. Qing, and D. Zhang, "Palm-print classification by global features," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 2, pp. 370–378, Mar. 2013.

[4] Y. Gong, D. Zhang, P. Shi, and J. Yan, "Handheld system design for dual-eye multispectral iris capture with one camera," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 5, pp. 1154–1166, Sep. 2013.

[5] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Toward unconstrained fingerprint recognition: A fully touchless 3-D system based on two views on the move," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 2, pp. 202–219, Feb. 2016.

[6] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial gummy fingers on fingerprint systems," in *Proc. SPIE*, vol. 4677. 2002, pp. 275–289.

[7] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[8] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[9] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (EEG) based authentication," in *Proc. 5th Int. IEEE EMBS Conf. Neural Eng.*, Apr./May 2011, pp. 442–445.

[10] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–752, Apr. 2007.

[11] R. Palaniappan, "Two-stage biometric authentication method using thought activity brain waves," *Int. J. Neural Syst.*, vol. 18, no. 1, pp. 59–66, 2008.

[12] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "A new multi-level approach to EEG based human authentication using eye blinking," *Pattern Recognit. Lett.*, vol. 82, pp. 216–225, Oct. 2015.

[13] S. Wang, Y.-C. Yu, and L. Gabel, "An application of the P300 event-related potential as a biometric," in *Proc. 40th IEEE Annu. Northeast Bioeng. Conf. (NEBEC)*, Apr. 2014, pp. 1–2.

[14] J. R. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. Oxford, U.K.: Oxford Univ. Press, 2012.

[15] M. Yu, N. Kaongoen, and S. Jo, "P300-BCI-based authentication system," in *Proc. 4th IEEE Int. Win. Conf. Brain Comput. Interface (BCI)*, Feb. 2016, pp. 1–4.

[16] I. Martinovic *et al.*, "On the feasibility of side-channel attacks with brain-computer interfaces," in *Proc. 21st USENIX Security Symp.*, 2012, pp. 143–158.

[17] B. Choi and S. Jo, "A low-cost EEG system-based hybrid brain-computer interface for humanoid robot navigation and recognition," *PloS ONE*, vol. 8, no. 9, 2013, Art. no. e74583.

[18] N. Kaongoen and S. Jo, "A novel hybrid auditory BCI paradigm combining ASSR and P300," *J. Neurosci. Methods*, vol. 279, pp. 44–51, Mar. 2017.

[19] A. Campbell *et al.*, "NeuroPhone: Brain-mobile phone interface using a wireless EEG headset," in *Proc. 2nd ACM SIGCOMM Workshop Netw. Syst. Appl. Mobile Handhelds*, 2010, pp. 3–8.

[20] J. Jin, I. Daly, Y. Zhang, X. Wang, and A. Cichocki, "An optimized ERP brain–computer interface based on facial expression changes," *J. Neural Eng.*, vol. 11, no. 3, 2014, Art. no. 036004.

[21] T. Kaufmann, S. M. Schulz, C. Grünzinger, and A. Kübler, "Flashing characters with famous faces improves ERP-based brain–computer interface performance," *J. Neural Eng.*, vol. 8, no. 5, 2011, Art. no. 056016.

[22] J. Jin *et al.*, "The changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement," *PloS ONE*, vol. 7, no. 11, 2012, Art. no. e49688.

[23] OpenBCI. (2016). *Open Source Biosensing Tools (EEG, EMG, EKG, and More)*. Accessed: Jun. 15, 2016. [Online]. Available: http://www.OpenBCI.com

[24] Florida Research Instruments. (2016). *Florida Research Instruments*. Accessed: Jun. 15, 2016. [Online]. Available: http:// floridaresearchinstruments.com

[25] U. Hoffmann, J. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain–computer interface for disabled subjects," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 115–125, 2008.

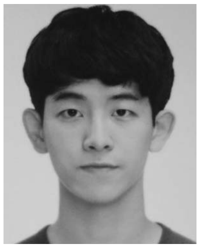
[26] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.

[27] G. Xu, F. Shen, and J. Zhao, "The effect of methods addressing the class imbalance problem on P300 detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2013, pp. 1–5.

[28] BZ. Allison and C. Neuper, "Could anyone use a BCI?" in *Brain-Computer Interfaces*. London, U.K.: Springer, 2010, pp. 35–54.

Netiwit Kaongoen was born in Phetchaburi, Thailand. He received the B.S. degree in bio and brain engineering and the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the School of Computing.

His current research interests include machine learning, brain–computer interfaces, cognitive neuroscience, and wearable computing.



Moonwon Yu received the B.S. degree in computer science from Kyungpook National University, Daegu, South Korea, in 2014 and the M.S. degree in computer science from the Korea Advance Institute of Science, Daejeon, South Korea, in 2016.

His current research interests include machine learning, artificial intelligence, and intelligent systems.

Sungho Jo (M'09) received the B.S. degree from the School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea, in 1999, and the M.S. degree in mechanical engineering and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2001 and 2006, respectively.

He was with the Computer Science and Artificial Intelligence Laboratory and the Laboratory for Information Decision and Systems, MIT. From 2006 to 2007, he was a Post-Doctoral Researcher with the MIT Media Lab. Since 2007, he has been with the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, where he is currently an Associate Professor. His current research interests include brain–machine interface, muscle–computer interface, and wearable computing.