I Think, Therefore I Am: Usability and Security of Authentication Using Brainwaves *

John Chuang[1], Hamilton Nguyen[2], Charles Wang[2], and Benjamin Johnson[3]

[1]School of Information, UC Berkeley
[2]Department of EECS, UC Berkeley
[3]Department of Mathematics, UC Berkeley
chuang@ischool.berkeley.edu
hamiltonnguyen@berkeley.edu
charleswang@berkeley.edu
benjamin@math.berkeley.edu

Abstract. With the embedding of EEG (electro-encephalography) sensors in wireless headsets and other consumer electronics, authenticating users based on their brainwave signals has become a realistic possibility. We undertake an experimental study of the usability and performance of user authentication using consumer-grade EEG sensor technology. By choosing custom tasks and custom acceptance thresholds for each subject, we can achieve 99% authentication accuracy using single-channel EEG signals, which is on par with previous research employing multi-channel EEG signals using clinical-grade devices. In addition to the usability improvement offered by the single-channel dry-contact EEG sensor, we also study the usability of different classes of mental tasks. We find that subjects have little difficulty recalling chosen "pass-thoughts" (e.g., their previously selected song to sing in their mind). They also have different preferences for tasks based on the perceived difficulty and enjoyability of the tasks. These results can inform the design of authentication systems that guide users in choosing tasks that are both usable and secure.

Keywords: pass-thoughts, EEG, authentication, usability

1 Introduction

Advances in EEG (electro-encephalography) bio-sensor technologies have opened up brainwave research and application development at an unprecedented level in recent years. Traditionally, EEG data capture has been performed in clinical settings using invasive probes under the skull or wet-gel electrodes arrayed over the scalp. Now, similar data can be collected using consumer-grade non-invasive dry-contact sensors built into audio headsets and other consumer electronics. This

* This research was supported in part by the National Science Foundation under award CCF-0424422 (TRUST).

opens up immense possibilities for using brainwave signals in different application domains. Originally limited to neuroscience research and clinical treatment of neurological diseases, EEG technologies are now being deployed for education, training, entertainment, and other ubiquitous computing applications.

Given the growing commercial availability of this technology, an important research agenda is to develop and evaluate different practical methods for regular users to apply their own brainwave data, in everyday (i.e., non-laboratory) settings, for different computer-based applications. In this work, we take a first step by focusing on the problem of user authentication using brainwaves. We propose and evaluate different classes of mental and/or motor tasks that users may perform while wearing a headset with EEG sensors. In addition to collecting EEG data from human subjects as they performed these tasks, we also collected experimental and questionnaire data to measure the usability of the tasks. Taken together, we compare the performance of different mental/motor tasks using metrics for signal similarity, authentication accuracy, task difficulty, task enjoyability, and task repeatability.

We make a significant departure from previous EEG-based authentication studies by studying the efficacy of single-channel as opposed to multi-channel EEG signals. Modern clinical EEG systems employ dense arrays of electrodes to provide 32, 64, 128, and 256 channels of EEG data. In contrast, for our experimental study, we use a consumer-grade headset that provides a single-channel EEG signal. Specifically, the Neurosky MindSet [1] places a single dry-contact sensor over the left frontal lobe region of the brain (Figures 1 and 2). Other than the EEG sensor, the headset is indistinguishable from a conventional Bluetooth headset for use with mobile phones, music players, and other computing devices. The headset can be purchased in the market for approximately $100.

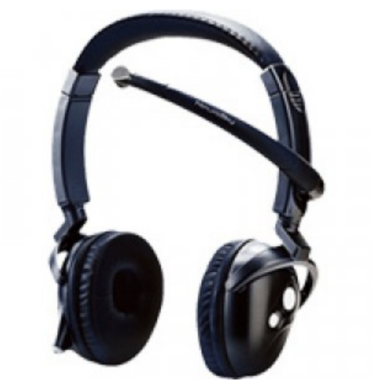


Fig. 1: EEG Headset Used in the Study: Neurosky MindSet

The headset form factor and the non-intrusiveness of the sensor imply a significant lowering of the usability barrier for EEG-based authentication. On the other hand, does the switch from multi-channel to single-channel signals lead to information loss that may render EEG-based authentication infeasible? This is a key motivating question of our study.

Our first key finding is that single-channel EEG signals do exhibit patterns that are subject-specific. Using standard measures of statistical similarity, we find higher signal similarity within subjects than across subjects. This is true across different mental tasks performed by the subjects; and it is true even for the brainwave signals of the same subjects that were collected over different experimental sessions on different days.

Our second key finding is that single-channel EEG authentication can be just as accurate as multi-channel EEG authentication. Leveraging our first finding, we propose and evaluate a suite of threshold-based authentication protocols that makes accept/reject decisions based on statistical similarities of signals. By combining the use of custom tasks and custom thresholds for each user, we can reduce false error rates down to the 1% level, which is comparable to the error rates achieved with multi-channel EEG signals.

Our third key finding is that neither signal similarity nor authentication performance are significantly affected by the categories of mental tasks performed by the subjects. In particular, personalized mental tasks (e.g., sing their favorite song silently, focus on their personal pass-thought) do not produce higher signal similarity or authentication accuracy over mental tasks that are common to all subjects (e.g., close eyes and focus on breathing).

On the other hand, as our fourth key finding, we find that the different categories of mental tasks score very differently in terms of user-perceived difficulty and enjoyability. When asked to choose a mental task that they would be willing to repeat on a daily basis, different subjects assign different weights to difficulty and enjoyability in making their choice. However, recall rates are consistently high for those mental tasks that require the subjects to remember their chosen secrets across sessions.

Taken together, these findings suggest that designers of EEG-based authentication systems do not have to make a hard choice between security and usability. The authentication system should be designed to allow users to experiment with different categories of mental tasks, so that each user repeats a customized task – one that they find easy and enjoyable, but that is also capable of producing high authentication accuracy.

2 Related Work

This research draws upon foundations and recent advances in multiple disciplines, ranging from neuroscience, human-computer interaction, computer security, signal processing, and machine learning. To the best of our knowledge, this work is the first experimental study of the usability design of brainwave-based authentication.

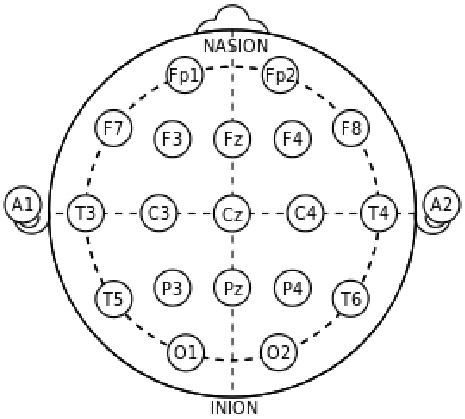


Fig. 2: Electrode placements for the International 10-20 Standard. The placement of the Neurosky Mindset electrode corresponds to the Frontal Polar 1 (Fp1) location.

2.1 Brainwave-based Authentication

The use of brainwave signals for user authentication has received widespread attention in recent years. Thorpe et al. motivate and outline the design of a "pass-thoughts" system [17]. By thinking of a pass-thought rather than typing in a password, this method of authentication promises numerous security advantages, including the resistance to dictionary attacks and shoulder-surfing.

A number of researchers have separately established the feasibility of using EEG signals to classify and/or authenticate users. With a focus on accuracy, they apply a range of statistical, signal processing, and machine learning techniques on multi-channel EEG signals. Poulos et al. use an artificial neural network to classify 4 subjects based on their EEG signals [16]. Marcel and Millan employ gaussian mixture model and maximum a-posteriori model for authentication with 9 subjects [12]. Palaniappan achieved 100% accuracy in classifying 5 subjects using a linear discriminant classifier [14], as well as zero False Acceptance Rate (FAR) and zero False Rejection Rate (FRR) using a two-stage threshold-based authentication process [15]. In each of these studies, the EEG data are captured using clinical-grade multi-channel sensors. More recently, Ashby et al. achieved 100% authentication accuracy with 5 subjects using consumer-grade multi-channel sensors [3]. In each of these studies, all the subjects performed identical tasks, ranging from baseline relaxation to imaginary motor movement, visualization, and solving math problems. None of these studies addresses task personalization or system usability.

2.2 Usability of Novel Authentication Systems

It is well understood that authentication systems must strike a balance between security and usability. Many security solutions fail not because of any flaw in the underlying technical design, but because of difficulties faced by humans in using the system in real-world settings as intended by the system designers. For example, users may find it difficult to remember one different password for each account they own, and resort to writing down the passwords on paper, thereby introducing new vulnerabilities to the system.

Such considerations underpin the development of graphical passwords as usable alternatives to text-based passwords [4]. In systems such as Draw-A-Secret [11], Deja Vu [7] and Passfaces [2], users authenticate themselves via recalling or recognizing images, rather than typing in a sequence of alphanumeric characters as in traditional password-based systems. A key usability metric for these systems is recall, i.e., the ability for users to remember their chosen secrets (e.g., images, faces) over different experimental sessions that are separated by periods of days or weeks. Usability studies demonstrate far higher recall rates for graphical passwords than for text passwords [7, 5]. In our experiment, we also investigate the ability for users to recall their chosen pass-thoughts across different sessions.

More generally, the different approaches to biometrics, including fingerprinting, iris scanning, facial recognition, voice recognition, each introduce different usability challenges and opportunities [6]. With the embedding of EEG and other bio-sensors into mobile phones, headsets, wearable computing devices, and other consumer electronics, the collection of brainwave signals for authentication and other purposes may become more natural and less intrusive than the collection of fingerprints, voice samples, and other biometric signals.

2.3 EEG and HCI

Research in brain-computer interface (BCI) has established the feasibility of using EEG signals to control computers and other devices. BCI systems can reliably evoke and measure event-related potentials (ERP) such as the P300, and use them to spell words and move computer cursors based on a user's intent [8–10, 13]. This proves very valuable in restoring the ability to communicate for patients suffering from the locked-in syndrome and other neurological diseases, and can be generalized to healthy users as well. While our work does not seek to infer user intent from their EEG signals, our choice of user tasks involving external stimuli are informed by the efficacy of eliciting and capturing these event-related potentials.

3 Experiment

3.1 Overview

Our research involved human subjects, and our experimental procedures were approved by an Institutional Review Board. We recruited a total of 15 subjects to participate in our study, all of whom were UC Berkeley undergraduate or graduate students. Each subject met with two investigators in a quiet, closed-room setting for two 40-50 minute sessions on separate days. We briefed subjects on the objective of the study, fitted them with a Neurosky MindSet headset, and provided instructions for completing each of seven tasks. As the subjects performed each task we monitored and recorded their brainwave signals.

3.2 Tasks

The following tasks were repeated five times in each session for each subject.

Breathing Task (breathing) Subjects close their eyes and focus on their breathing for 10 seconds.

Simulated Finger Movement (finger) Subjects imagine in their mind that they are moving their right index finger up and down in sync with breathing, without actually moving their finger, for 10 seconds.

Sports Task (sport) Subjects select a specific repetitive motion from a sport of their choosing. They then imagine moving their body muscles to perform the motion, for 10 seconds.

Song/Passage Recitation Task (song) Subjects imagine that they are singing a song or reciting a passage for 10 seconds without making any noise.

Eye and Audio Tone Task (audio) Subjects close their eyes and listen for an audio tone. After 5 seconds, the tone plays; upon hearing the tone, the subjects open their eyes and stare at a dot on a piece of paper in front of them for an additional 5 seconds.

Object Counting Task (color) Subjects are asked to choose one of four colors – red, green, blue, or yellow. They are then shown on a computer screen a sequence of six images. Each image contains a 5x6 grid of colored boxes. As each grid appears, subjects count, silently in their mind, the number of boxes corresponding to their chosen color. A new grid appears after each 5 seconds. The task continues 6 rounds for a total of 30 seconds.

Pass-thought Task (pass) Subjects are asked to choose their own pass-thought. A pass-thought is like a password; however, instead of choosing a sequence of letters and numbers, one chooses a mental thought. When subjects are instructed to begin, they focus on their pass-thought for 10 seconds.

3.3 Questionnaire

In addition to the brainwave data, we also asked subjects a series of survey questions. At the end of each session, we asked the subjects to select the one task (out of seven) that they would be most willing to repeat every day. After subjects completed both sessions, we asked them to rate each of the tasks according to the following binary choices: (i) difficult or easy, and (ii) enjoyable or boring.

3.4 Brainwave Data

As subjects completed each task, we recorded their raw EEG data on a computer. The data was transmitted via a bluetooth network connection from the headset to the computer. The raw data includes single-channel EEG signals in both the time and frequency domains. We specifically use the power spectrum data, a two-dimensional matrix which gives the magnitude of the signal for every frequency component at every point in time. With 15 subjects repeating seven tasks, five times per session, and two sessions per subject, we have a total of 1050 brainwave data samples.

3.5 Data Preprocessing

Before performing any analysis on the brainwave data, we first pre-process the power spectrum data to compress the samples. In the temporal dimension, we extract only the middle five seconds out of the total ten seconds of each recorded signal (the exception is the *color* task, for which we chose a five-second section corresponding to a specific image). In the frequency dimension, we extract only the data corresponding to the alpha wave (8-12 Hz) and the beta wave (12-30 Hz) ranges of the signals. We apply our analysis to both ranges.

The second step in our data preparation is to take this two dimensional signal and compress it into a one dimensional signal. Our chosen compression method flattens the signal in the time dimension – specifically, for each frequency component, we compute the median magnitude corresponding to that frequency component over all time. The end result is a one-dimensional column vector with one entry for each measured frequency. This column vector representation is how brainwave samples are stored and manipulated within the authentication system.

4 Data Analysis

After collecting and processing the brainwave data, we begin evaluating the effectiveness of the signals in the context of authentication. This problem requires us to distinguish the signals among different subjects.

We begin by quantifying the similarity between two signals u and v as the cosine similarity of the vector representation of the signals, given by the equation:

$$\text{similarity}(u, v) = \frac{u \cdot v}{\|u\|\|v\|}.$$

Similarity gives a value between 0 and 1, where a similarity of 1 would indicate a perfect match.

We next define two additional notions related to similarity – self-similarity and cross-similarity. Self-similarity refers to the similarity of signals within a single subject, while cross-similarity refers to the similarity of signals between different subjects. Our hypothesis is that self-similarity should be consistently greater than cross-similarity for all subjects, in all tasks. If this is true, we will be able to leverage this difference in our authentication system.

For a fixed task t and given subject s, we define the self-similarity of s in t to be the average of the similarity of every possible pair of samples belonging to s. Likewise, for a fixed task t and given subject s, we define the cross-similarity of s in t to be the average of the similarity of every possible pair for which one sample in the pair belongs to s and the other sample does not belong to s.

Table 1 displays the results of testing our similarity metric. For a given subject, we compute his or her self- and cross-similarity for every task, and then take the average of these values. The final average is the number displayed under the Self and Cross columns. Lastly, we look at the relative difference between self- and cross-similarity for each subject rather than the absolute difference. The last column corresponds to the percent difference between the Self and Cross columns.

From these results, we can make a few observations. First, self-similarity is higher than cross-similarity for all subjects, which is an important pre-requisite in using this metric in our authentication system. Second, there is noticeable variation in percent difference between the 15 subjects. This second result will be used in improving our protocol.

Next, Table 2 gives an alternative visualization of our results. For a given task, we compute the self- and cross-similarity of each subject, and then take the average over all subjects. This gives similarity values associated with tasks rather than subjects. Again, we can see that self-similarity is higher than cross-similarity in all cases. Interestingly, we can observe that the variance in difference in Table 1 is higher than the variance in difference in Table 2. This suggests that the similarity measure has greater variation between subjects than between tasks.

Table 1: Similarity Comparison of Subjects

	Self	**Cross**	**Percent**
Subject	**Similarity**	**Similarity**	**Difference**
subject 0	0.7207	0.6653	7.99%
subject 1	0.7268	0.6745	7.46%
subject 2	0.7014	0.6602	6.05%
subject 3	0.7577	0.6397	16.89%
subject 4	0.7232	0.6617	8.88%
subject 5	0.6771	0.6702	1.02%
subject 6	0.7147	0.6264	13.17%
subject 7	0.7253	0.6817	6.20%
subject 8	0.7368	0.6828	7.61%
subject 9	0.6941	0.6435	7.57%
subject 10	0.7161	0.6847	4.48%
subject 11	0.7142	0.6816	4.67%
subject 12	0.711	0.6817	4.21%
subject 13	0.7028	0.6106	14.04%
subject 14	0.7099	0.6702	5.75%

Table 2: Similarity Comparison of Tasks

	Self	**Cross**	**Percent**
Task	**Similarity**	**Similarity**	**Difference**
breathing	0.7304	0.6834	6.65%
finger	0.7282	0.6567	10.33%
sport	0.7144	0.676	5.52%
song	0.7013	0.6498	7.62%
audio	0.7283	0.6637	9.28%
color	0.6664	0.599	10.65%
pass	0.6931	0.632	9.22%

5 Authentication

5.1 Problem Definition

The authentication problem is also referred to as the user verification problem. Given an (identity, sample) pair, the authentication system must determine if the sample provides a legitimate match to the identity.

Authentication systems make two types of errors: False Acceptance (FA) errors occur when the system accepts an impostor, while False Rejection (FR) errors occur when the system rejects an authorized user. The performance of an authentication system can thus be measured in terms of its False Acceptance Rate (FAR) and False Rejection Rate (FRR). The two error measures are often merged to form the Half Total Error Rate (HTER), defined as:

$$HTER = (FAR + FRR)/2.$$

5.2 Testing Schema

Before discussing the implemented authentication protocols themselves, we briefly describe our testing schema used to evaluate the performance of the protocols. Recall that for each task and subject we collected and processed 10 brainwave samples. Our testing schema randomly selects 5 of these samples (for each task and user) to train the authentication protocol. The remaining samples are used to test the protocol.

Evaluating FRR To assess false rejection we may focus our attention on a single user at a time. Given a specific task, each user has only 5 samples in the testing set for that task, and our testing schema runs the relevant authentication protocol on each of them along with the user's correct identity. If the protocol were to work perfectly it would always accept these (user, sample) pairs. The FRR is computed as the average percentage of such tests that do not accept, taken over all matching pairs of users and samples in the test set.

Evaluating FAR To assess false acceptance we must focus on many users at a time. Indeed as there is only one legitimate user but many potential impostors, there are many more opportunities for false acceptance than for false rejection. Given a specific task and user, our testing schema randomly selects 5 samples that do not match the user, and runs the relevant authentication protocol on this set of false (user, sample) pairs. If the protocol were to work perfectly it would always reject these pairs, and the FAR is computed as the average percentage of such tests that incorrectly accept.

5.3 Protocols and Results

Baseline Protocol Our baseline protocol will also be referred to as the Common Task Common Threshold protocol. In this system, all brainwave samples correspond to a single, fixed task. We then choose a common threshold T to be used for all subjects.

The core authentication mechanism is as follows: a user provides as input his claimed identity and brainwave sample. We compute the value $selfSim$ to be the average similarity between the given sample and all 5 samples known to belong to the user. We then randomly select a set of 5 samples such that none of the samples in this set belong to the user. Next, we compute the value $crossSim$ to be the average similarity between the given input sample and the samples in this new set. Finally, if the percent difference between $selfSim$ and $crossSim$ is greater than or equal to T, we accept the authentication attempt. If not, we reject it.

Table 3 shows the result of testing the baseline protocol for each of the tasks. Although the protocol performs better than random guessing, it is still far from practically usable. At best, the HTER is at .322 for the *audio* task. We also observe that FAR is lower than FRR for every task (and for most tasks, many times lower), which implies the current protocol is more effective at determining impostors than confirming legitimate users. In the following sections, we explore improvements over the baseline protocol.

Table 3: Authentication with Common Thresholds

Task	**FAR**	**FRR**	**HTER**
breathing	0.156	0.578	0.367
finger	0.044	0.733	0.389
sport	0.089	0.644	0.367
song	0.155	0.578	0.367
audio	0.244	0.400	0.322
color	0.244	0.622	0.433
pass	0.356	0.400	0.378

Table 4: Authentication with Customized Thresholds

Task	**FAR**	**FRR**	**HTER**
breathing	0.000	0.280	0.140
finger	0.067	0.120	0.093
sport	0.027	0.187	0.107
song	0.000	0.093	0.047
audio	0.027	0.147	0.087
color	0.120	0.440	0.280
pass	0.000	0.120	0.060
customized	0.000	0.022	0.011

Customized Threshold Our first improvement over the baseline is the Common Task Customized Threshold protocol. This new protocol is nearly the same as the baseline except for one key difference – rather than comparing against a common threshold T, we compare against T_i, a customized threshold optimized specifically for user i.

The first seven rows of Table 4 show the results of testing the performance of this new protocol for each of the tasks. With customized thresholds, we were able to decrease FRR significantly for every task, and in almost every case, we did not sacrifice performance with regards to FAR – the lone exception is the $finger$ task, for which FAR actually increased when customized thresholds were implemented. Overall however, the HTER of the $finger$ task decreased as well.

Further, we were able to achieve a reasonably high success rate for nearly all tasks. Put another way, these results do not suggest that there is one particular kind of task that is definitively most effective for authentication.

Customized Task Customized Threshold Our final protocol is the Customized Task Customized Threshold protocol. In the previous two protocols, the chosen task was fixed for all subjects. We add an additional step of precomputation in which we determine for each subject, the optimal task to maximize the difference between self and cross similarity for the subject. Then, within that task we determine the optimal threshold specific for that subject, as above.

The last row in Table 4 shows the result of using customized tasks. This version of the protocol outperforms every instance of using common tasks, achieving an HTER of 1.1%. The success of the customized task protocol further reinforces our belief that there does not exist one single task that is the best to use for authentication.

Additionally, we can remove tasks one-by-one from the pool of tasks considered by the protocol and observe how this affects performance. In one instance, we were able to reduce the pool of tasks to only two – specifically *breathing* and *audio* – and still maintain the same HTER of 1.1% as when all seven tasks are used.

6 User Identification

We next consider the more challenging problem of user identification, i.e., given a brainwave signal, can we identify the user to which the signal belongs. This corresponds to the classification problem in machine learning, and we apply standard classification techniques to our data. As in our approach to authentication, we first prepare a truncated signature for each trial by restricting to alpha wave and beta wave frequencies, and averaging across the middle portion of the time domain.

Our testing schema is then as follows: we select one trial signature to be a testing sample. The remainder of the trial signatures are treated as training samples, i.e., we assume the subject identities of these signatures are known. We ask our classifier to classify the testing sample, and record whether the classifier identified the correct subject. This process is repeated for every trial signature.

Our classifier is a basic adaptation the K-Nearest Neighbors (KNN) algorithm for coloring graphs. Given a complete graph with distances between each node and with all but one node colored, the KNN algorithm colors the uncolored node with the most common color among its K nearest neighbors. If there is a tie among colors in the nearest neighbor set, we restrict to nodes having those tied colors, and run the algorithm again with K decremented. Any ties remaining when $K = 1$ are resolved by a fair coin flip. Our adaptation of this algorithm has trial signatures as nodes, subject identities as colors, and Cosine Similarity as the distance metric.

Figure 3 summarizes the classification success rates for $K = 5$. The classifier generally does two to three times better than random guessing. (Since there are 15 colors, random guessing has a classification success probability of $\frac{1}{15} \approx 6.7\%$.) The *audio*, *sport*, and *color* tasks have the best overall classification rates. For example, the classifier can correctly identify a user 22% of the time based on EEG samples from the *audio* task. This corresponds to a 3.3x improvement over random guessing. Nonetheless, a 22% success rate still falls far below levels acceptable for practical user identification systems.

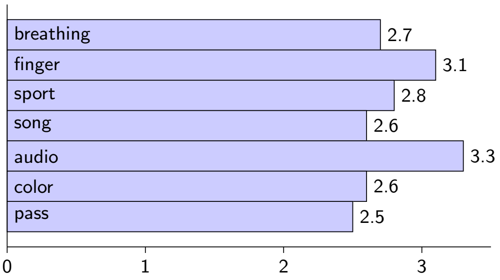
The reason for the discrepancy in performance between user authentication and user identification is instructional. For user authentication, we can pick custom tasks that provide the highest authentication accuracy for each subject. For user identification, on the other hand, knowledge of which task was performed for a given EEG sample does not help in the classification at all.

7 Usability

There are two dimensions of usability to consider: the usability of the EEG hardware, and the usability of the mental tasks.

In terms of hardware, a single-channel EEG sensor in the form of a dry-contact electrode integrated with a wireless headset is much less intrusive than an array of electrodes that must be carefully placed over the scalp. Having established that single-channel EEG signals collected with consumer-grade EEG

Fig. 3: Classification Performance (random-guessing = 1.0)



sensors over a range of mental tasks can provide the same level of authentication accuracy as multi-channel EEG signals collected with clinical-grade EEG sensors, we can posit that the usability vs. security tradeoff is now tipping in favor of the consumer-grade single-channel approach.

Let us turn to the usability of the mental tasks. At the conclusion of the second experimental session, each of the fifteen subjects was asked in a questionnaire to rate each of the seven tasks as either "difficult" or "easy", and as "boring" or "enjoyable". The responses are summarized in the first three columns of Table 5. For example, seven of fifteen subjects found the *pass* task to be difficult to perform, because their chosen pass-thoughts involve feelings or events that proved hard to repeat on a consistent basis. Similarly, seven of fifteen subjects found the *sport* task to be difficult to perform, because they found it unnatural to imagine the movement of their muscles without actually moving them. On the other hand, all fifteen subjects found the *breathing*, *audio*, and *color* tasks to be easy to perform.

Eight of fifteen subjects rated the *finger* task as boring. Presumably, the task is monotonous just as it is easy. On the other hand, twelve of fifteen subjects rated the *breathing*, *sport*, *song*, and *color* tasks as enjoyable.

At the conclusion of both the first and second experimental sessions, the questionnaire also asked the subjects to choose one task that they would most like to repeat on a daily basis. The responses are summarized in the last column of Table 5.

We can see that the *finger* task, rated boring by more than half of the subjects, was not chosen at all. The *sport* task, rated difficult by almost half of the subjects, received the next fewest votes. On the other hand, the *color* and *breathing* tasks received overall the most repeatability votes. These two tasks are the least boring and least difficult tasks, as evaluated by the subjects.

Two of the seven tasks require the subjects to respond to external stimuli – an audio tone for the *audio* task, and a sequence of images for the *color* task. Both tasks were perceived to be easy and likely candidates for daily repetition.

Four of the seven tasks provide the subjects an opportunity to choose their own secret: *sport*, *song*, *color*, and *pass*. In contrast, the other three tasks, *breathing*, *finger*, and *audio*, do not involve a personal secret. We do not observe

Table 5: Usability Comparison of Tasks

Task	Was Difficult	Was Boring	Would Repeat
breathing	0/15	3/15	7/30
finger	3/15	8/15	0/30
sport	7/15	3/15	1/30
song	4/15	3/15	5/30
audio	0/15	4/15	4/30
color	0/15	3/15	9/30
pass	7/15	6/15	4/30

any relationship between the utilization of a secret and the difficulty, enjoyability, or repeatability of a task.

During the second experimental session, we tested each subject on their ability to recall their chosen secrets for the *sport*, *song*, *color*, and *pass* tasks. As seen in Table 6, the subjects had no difficulty in recalling their personalized sport, song, and pass-thought choices. One of the fifteen subjects could not recall the color he chose from the previous session. This suggests the possibility that users are better able to remember secrets that they come up with themselves, than secrets that they select from a menu of discrete choices.

An open question is whether the changing of a chosen secret, as part of a user-initiated password change routine, may affect the authentication performance or even the usability of the task.

Table 6: Recall Rate of Tasks

Task	Recall Rate
song	15/15
sport	15/15
color	14/15
pass	15/15

8 Conclusion and Future Work

In this paper, we study the usability and performance of brainwave-based authentication. Motivated by the trend of low-cost EEG sensors embedded in various consumer electronic devices, we conduct an experimental study to capture brainwave signals from human subjects using consumer-grade EEG headsets in a non-clinical environment. We design a number of different mental tasks for the subjects to perform, and evaluate the usability of the tasks based on their difficulty, enjoyability, and repeatability.

We find that brainwave signals, even those collected using low-cost non-intrusive EEG sensors in everyday settings, can be used to authenticate users with high degrees of accuracy. We show it is possible to compensate for the lower fidelity of single-channel EEG signals by intelligently matching signal similarity thresholds and customized tasks to each user. This means that we can now bypass the usability challenges associated with conventional EEG systems designed for clinical applications.

Different mental tasks also vary in their usability. Subjects will not opt for repeating tasks that are perceived as either difficult or boring. Similar to the experience with graphical passwords, we find that pass-thoughts chosen by the subjects can be recalled by the subjects without much difficulty. In comparing the results of the usability analysis with the results of the authentication testing, we observe that there is no need to sacrifice usability for accuracy. It is possible to achieve accurate authentication with easy and enjoyable tasks.

There are a number of limitations of our study that point to interesting directions for future work.

We are able to maintain a high level of authentication accuracy with a subject pool that is 66% to 275% larger than those from previous studies [3, 12, 14–16], thus demonstrating the feasibility of authentication in a small population, e.g., a work group setting [14]. Nonetheless, it would still be valuable to investigate the scalability of the results to even larger populations.

Going beyond the small set of mental tasks evaluated in this study, a systematic exploration of additional categories of tasks would be of great value. From there, we can seek to gain a more complete understanding of which factors influence the usability and security performance of mental tasks.

While the primary focus of this paper is on user authentication, we also encountered the relative difficulty of accurate classification of users. It would be useful to ascertain whether the classification performance can be improved with other classification algorithms or with other mental tasks.

The robustness of brainwave-based authentication against impersonation attacks is an interesting problem. If an attacker gains knowledge of a target's customized task and chosen secret (e.g., the specific song or passage that a user repeats to herself), how easy or difficult is it for the attacker to fool the authentication system by performing the same customized task?

Finally, if the authentication system works by choosing customized tasks for each subject, the user enrollment process becomes an important design consideration. Today's users may balk at a 45 minute initiation process to set up their password, so the number and choice of mental task categories have to be carefully selected to optimize for both the duration of user enrollment and the accuracy of authentication.

References

1. Neurosky MindSet. http://www.neurosky.com/.
2. Passfaces. http://www.passfaces.com/.

3. C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein. Low-cost electroencephalogram (eeg) based authentication. In *Proceedings of 5th International IEEE EMBS Conference on Neural Engineering*, April 2011.
4. R. Biddle, S. Chiasson, and P. van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 44(4), 2011.
5. S. Brostoff and M. A. Sasse. Are passfaces more usable than passwords? a field trial investigation. In *Proceedings of HCI*, 2000.
6. L. Coventry. Usable biometrics. In L. Cranor and S. Garfinkel, editors, *Usability and Security*. 2005.
7. R. Dhamija and A. Perrig. Deja vu: a user study using images for authentication. In *Proceedings of the 9th conference on USENIX Security Symposium*, 2000.
8. E. Donchin, K. Spencer, and R. Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, 8(2):174–179, June 2000.
9. L. A. Farwell and E. Donchin. Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523, 1988.
10. T. Hinterberger, A. Kubler, J. Kaiser, N. Neumann, and N. Birbaumer. A brain-computer interface (bci) for the locked-in: comparison of different eeg classifications for the thought translation device. *Clinical Neurophysiology*, 114(3):416–425, March 2003.
11. I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin. The design and analysis of graphical passwords. In *Proceedings of 8th USENIX Security Symposium*, August 1999.
12. S. Marcel and J. del R. Millan. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), April 2007.
13. F. Nijboer, E. Sellers, J. Mellinger, M. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. Krusienski, T. Vaughan, J. Wolpaw, N. Birbaumer, and A. Kubler. A p300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119(8):1909–1916, 2008.
14. R. Palaniappan. Electroencephalogram signals from imagined activities: A novel biometric identifier for a small population. In *IDEAL 2006, LNCS 4224*, pages 604–611, 2006.
15. R. Palaniappan. Two-stage biometric authentication method using thought activity brain waves. *International Journal of Neural Systems*, 18(1):59–66, 2008.
16. M. Poulos, M. Rangoussi, N. Alexandris, and A. Evangelou. Person identification from the eeg using nonlinear signal classification. *Methods of Information in Medicine*, 2002.
17. J. Thorpe, P. van Oorschot, and A. Somayaji. Pass-thoughts: Authenticating with our minds. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, 2005.