Recognition based Authentication using BCI Technology

Usability & Evaluation Test Plan

Keerti Kosana

18 April 2019

Table of Contents

Executive Summary

Brain-Computer Interfaces (BCIs) are an emerging trend with vast applications in both medical and commercial fields. A BCI acquires, analyzes, and translates signals from the brain into commands that a computer can understand. Because BCIs do not utilize normal neuromuscular output pathways, they can be used in medicine to greatly improve the quality of life for individuals that have disabilities, or severe neuromuscular disorders [2]. By developing prosthetics that operate using EEG (electroencephalography) signals from the brain, these individuals could feel as though they never lost their arm or leg [1]. As stated earlier, BCI technology also has applications extending beyond those in healthcare and neuro-medical science.

Studies have begun in user authentication, gaming and entertainment, as well as smartphone-based applications [1] to incorporate BCI technology. Our research study will focus on evaluating usability and security for a recognition based authentication system that implements BCI technology. Specifically, we will explore the elicitation of P300-neurological signals that are activated in an Event Related Potential (ERP) response [6]. In our study, we design methods that will assess the usability and security of a system that does not necessarily have a functional backend. Our goal is to come up with testing methods that accurately and comprehensively address the concerns of security and privacy in our proposed authentication system.

Our proposed system will comprise of headphones having functionality similar to that of a traditional EEG headset. Current EEG headsets are uncomfortable to wear for users; the bulky design discourages users from fully utilizing BCI technology to its full potential. To address this issue, we suggest a simple hardware prototype that is easier to use, carry, and wear. This prototype is meant to detect the P300 signals in the brain that are unique to every individual, and stimulated when the user is presented a dataset that he/she could recognize. To successfully conduct our research, we identified three types of graphical images that will be used for authentication: abstract, object, and text images. We expect the P300 signal to be activated when a user recognizes images within a dataset, essentially creating a recognition based authentication system that incorporates BCI technology.

To evaluate our proposed system, we recruited students at Clemson University. After the students were briefed on the purpose and nature of our study, we provided each of them with three different password sets for every image type. After the participants were given time to familiarize themselves with an image set, they were tested to determine how many of the password images they recognized. The results were recorded based on the recognition capabilities of the participants with measures like a predetermined time on task, success/failure rate, and false-positive rate, all of which are discussed in detail in the sections below. Finally, we conclude this study by discussing the limitations we recognized along the way, and by suggesting potential improvements on the system and its evaluation methods.

Introduction

Brain Computer Interfaces are a rapidly growing technology with a broad range of applications from the medical fields to information technology. Initially, BCIs were a major source of interest because of their ability to provide assistance to individuals with neurological disabilities. With time, a shift towards implementing BCI technology towards gaming and non-medical applications took place. Today, this technology has developed to the point where it is being used for the sole purposes of comfort and an easier lifestyle.

Alternatively, authentication systems are incorporating BCI technology to utilize the individuality in human thought. Existing research studies [4, 5] explore the concept of P300-based BCI authentication systems. The P300 wave is a neurological signal that is active only in an Event Related Potential (ERP) response that is elicited upon detecting a target internal or external stimulus [6]. This wave only occurs if the user is actively in the task of detecting the targets. Traditionally, it is tracked using electroencephalography (EEG) headsets. In our study, we propose a prototype that is significantly simpler in design to detect P300 signals. Our study operates under the assumption that the headset is able to detect P300 signals upon ERP response elicitations. Accordingly, we design evaluation methods for authentication based on recognition to test the usability of our proposed technology. Our research [9] suggests that recognition happens in about 150 ms in an average human. Following this evidence, we create authentication tests based on recognition for three different types of image sets to compare and analyze the similarities and differences. The following sections discuss the methods, procedures, evaluation techniques, usability testing, and results in detail.

Methodology

For our study, we recruited twenty students at Clemson University to partake in the usability testing for evaluation of our system. The testing was done in an informal setting wherever the participants felt comfortable with the help of a simple PowerPoint to display each of the three image sets; each participant was tested one at a time. Before beginning our evaluation, each participant was provided with three sets of eight-image password sets - object, abstract, text - to familiarize himself/herself with. We chose the images for the users as users tend to choose weak or biased passwords when given the opportunity to choose [12]; we address the security concern of this predictability by assigning the password set. The object and abstract sets were chosen due to the amount of research behind them, including systems like PassFaces and DejaVu. The text as a mechanism was included after some feedback we received; this system was also inspired by CAPTCHA systems that are commonly used. Then, three slide decks are prepared, including images from the password set and other distractor images belonging to the same image type were presented one at a time for each image type. To clarify, an abstract image set, for example, would consist of twenty images including a predetermined number of password images from the abstract image password set only. We chose to issue a subset of a given static password set such that the login set was not static, which should confer some higher degree of security to the users. The images in the slide deck were timed strategically with a latency of 350 ms between each slide, which was 7 sec total. Existing research [10] as displayed in Figure 1 below shows that P300 signal occurs in a timeframe between 300 and 400 ms in humans ages 20-30. Hence, we decided on an average of 350 ms to allow for recognition and the P300 stimulus response to occur, which may also be considered to account for a small degree of signal refractory period. We had the users respond to indicate when they recognized an image and at the end, counted the total number of images they recognized. This allows us to see if the users successfully authenticate by meeting the arbitrary number of images required to login; we used 4 for these tests. We additionally changed up the slide decks such that there were less than 4 images from the password set in them. After shuffling the pictures before retesting, we ran the authentication test once more. These tests were meant to help us identify the rate of false positives in our mechanism. By collecting data on the number of images each participant recognized, we expected to quantitatively analyze the success, failure, and the false-positive rate of recognition.

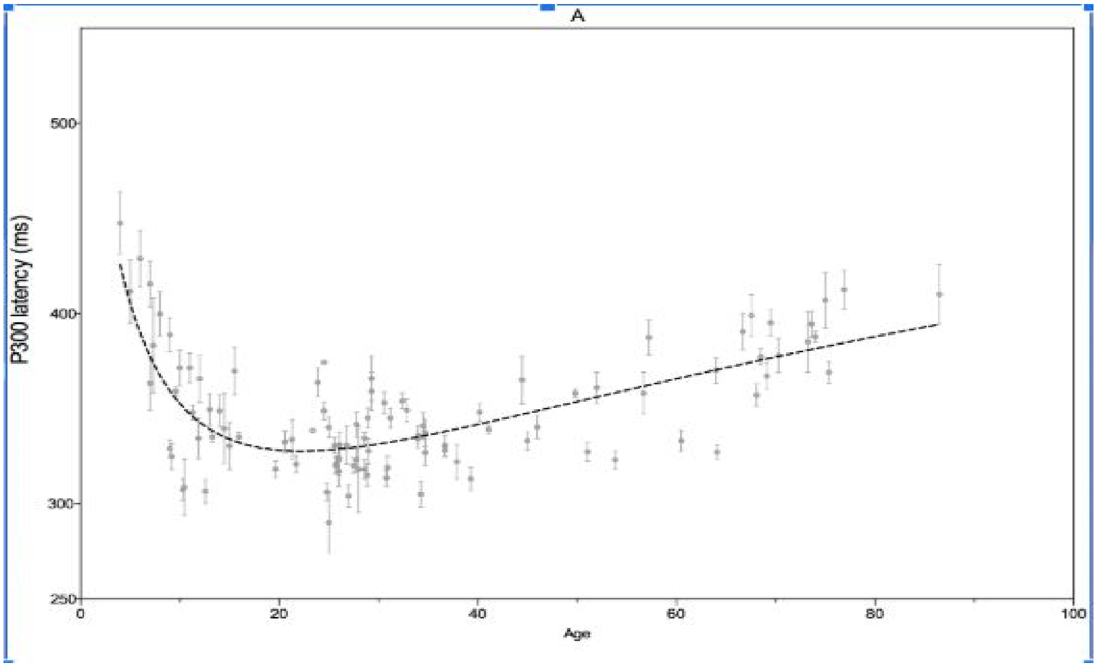


Figure 1. P300 latency and amplitude trajectories across the lifespan as obtained from the meta-analysis

User Research

Many current studies focus on implementing BCI technology for people with neurological, or other disorders pertaining to the mind. Since our proposed technology doesn’t have a functional backend, we were limited in diversifying our target user group. Meaning, we could not have conclusively tested for usability on participants with neurological disabilities. Some research studies we referenced [7] suggested that users in the age range between 18 and 29 were more likely to explore advanced technologies than other age groups. So, our study focused on college students who belong to a similar age range.

Aside from this study, as an authentication mechanism in general, we believe that the user group for the technology would be everyone who must login to a system, with a preference toward younger age ranges that may be more likely to adopt the technology. The way we designed our prototype does not work with individuals with visual impairments or severe short-term memory issues.

User Needs

Karen Renaud [8] stated that users tend to expect technology to be:

- Accessible – allow users to individualize technology based on their own needs/limitations
- Secure – minimize data breaches of personal and confidential information through shoulder surfing, or weak passwords.
- Fast and Repeatable

Our study focuses on recognition based authentication systems which allow for faster authentication because it takes approximately 150ms for recognition to occur, according to our research [9].

Currently, there is existing work on P300-based authentication as discussed extensively in [4] that aligns with our approach. The experimental setup and the concept of eliciting P300 signals in the brain through the use of external stimulus is similar to that of our research. However, a major difference between our works lies in the design of our evaluation methods. The existing study does not account for P300 signal response in the experimental setup. They use a monitor to display a randomly generated sequence of ten images, two of which are known to the user, with a latency of 100 ms between each image. The time on task is 1 second for their test and does not consider the evidence that P300 latency may vary from 250 ms to 750 ms from the onset of stimulus as identified in their proposal. In other words, the user is expected to finish the task within the blink of an eye, which is extremely fast. Our approach improves on their proposal by taking into account both recognition and P300 latencies.

Proposed Interface

Interface Design

The system we propose would be used much like other login systems that exist today, but the main difference is that there would be very little action taken by the user. In a traditional authentication system, once the user reaches their desired login interface, they would click on a login button, which could be time consuming. Our system proposes a design for authentication that would authenticate a user 'on the go' in as little as 7 seconds. This fast authentication is possible through the use of our headphones pictured in Figure 2 below. As the user is wearing the headphones, he/she then proceeds through a set of images that contains the required amount of pass-images (from the password set) to authenticate. Ideally, the headset wearable would read the P300 signal emitted from the user and the computer would recognize and count the number of signals emitted. From here the system would compare the number of P300 signals caught to the number expected. At the end of the comparison, the user would either be authenticated or denied by the system as a result of the caught signals matching the expected outcome or not.

Figure 2. Prototype of the headset used for testing

This is the system proposed as it is believed to fulfill the user needs of being accessible, secure as well as fast and repeatable. The system proposed is secure as we believe that it will reduce the risk of shoulder surfing since users do not explicitly acknowledge their correct password images. Also it is accessible to most users as all they need is the headset, although the system proposed would not work to its full potential for users with visual impairments. As stated before the system would also be relatively quick to proceed through as well as being repeatable.

Participants

All participants were students at Clemson University. Twenty participants were recruited and scheduled during the 2nd week of April, 2019 to perform the usability test. The group contacted and recruited participants from Clemson University (students). Participants were asked to select a date, time, and location that was convenient for them, which would revolve around their busy schedules.

Testing

Test sessions per participant would last approximated twenty minutes. Participants received printed instructions which were as follows:

1. You will be provided eight images from three different groupings that you are to spend up to five minutes familiarizing yourself with:
    a. Object images
    b. Abstract images
    c. Text images

2. After your five minute study session for each image set, you will be presented a slide deck of twenty images where each image is displayed for less than a second.
3. When you recognize an image in this presentation as one of the given eight, make a tally (participants were given the option of making a tally on paper provided, some opted to keep track with their hands and fingers).
4. Your examiner will ask how many images you recognized from each set after presentation.
5. You will receive two tests for each image set. Each test will contain a different combination and number of images from the password set.

After the last test was taken, the participants were asked the following overall questions:

- What the participant liked the most.
- What the participant liked the least.
- Recommendations for improvement.

Evaluation Tasks/Scenarios

Test participants attempted completion of the followings tests of image sets. Participants were expected to recognize:

- Four of the twenty images from Abstract Test 1
- Three of the twenty images from Abstract Test 2
- Four of the twenty images from Object Test 1
- Three of the twenty images from Object Test 2
- Four of the twenty text based images from Text Test 1
- Three of the twenty text based images from Text Test 2

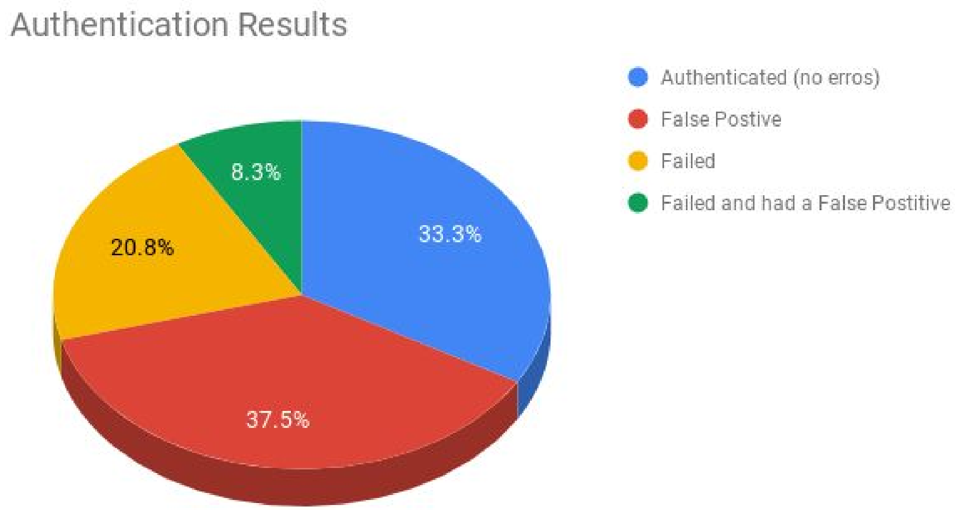
Our research group wanted to keep the number required to authenticate between three and four therefore the first test set contained four images required to authenticate and the second test set contained three images. For the four images to authenticate, the number was chosen as a means to randomize what pictures were in the 20 images when shown to the user. This was to reduce the risk of shoulder surfing from any malicious agent acting to determine what images were in the user's set of pass-images. The other number of three was chosen arbitrarily. It did not specifically have to be three, it just had to be less than the number used for authentication as the goal of the tests including three pass-images was for determining false positives.

Results

The following section gives a brief overview of the data collected followed by a categorical analysis of results obtained from our usability testing based on each image type. Also included is some feedback provided from participants of the tests. Feedback generally consisted of which image sets were preferred and easier to work with.

Summary of Data

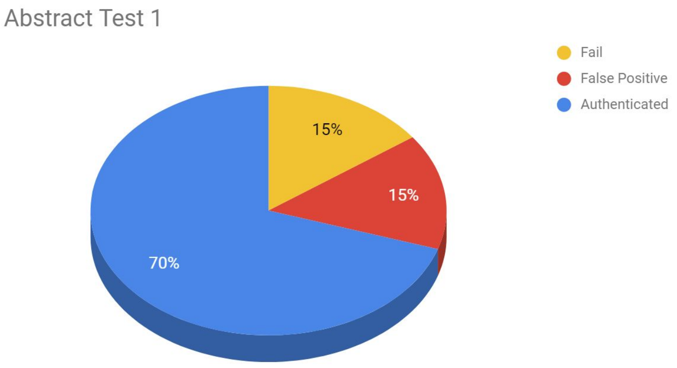
Eight of the twenty participants completed the test and were authenticated without any errors or false positives. Nine of the twenty participants had a false positive (recognized/counted an image not part of the eight image set). Five of the twenty participants failed to recognize the correct number of images. And finally, two of the twenty participants had both a failure to recognize, and a false positive from the tests. The following chart displays these data points.

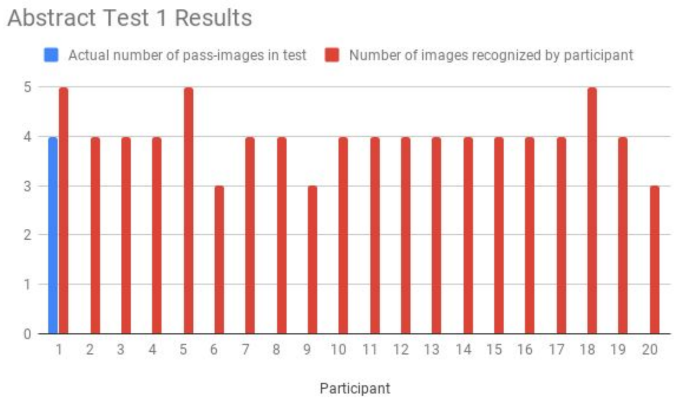


Results were collected and converted into column charts and pie charts displayed below for each test. You will note that the blue line in each column chart indicates the actual number of images from the original set of eight that the participant should have recognized. If the participants recognized more than the actual number of images, that would indicate a false positive. If participants recognized less than the actual number of images, that indicated a failure to authenticate.

Quantitative Analysis

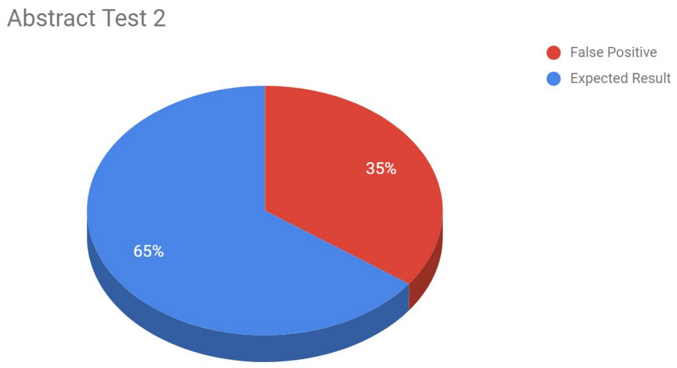
Abstract Test 1

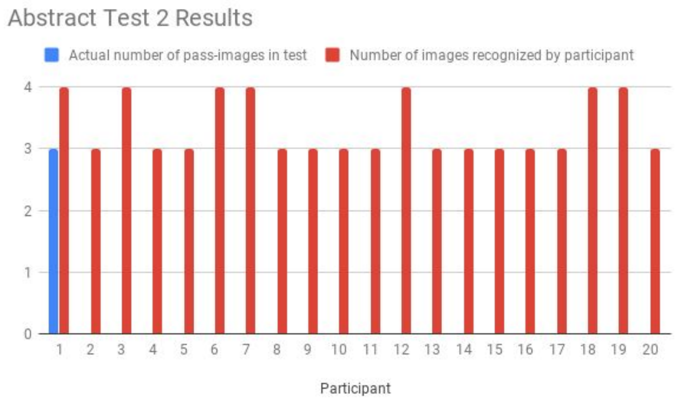




In the first test conducted, as can be seen on the right chart, there was a majority of 14 participants who succeeded in recognizing four images. There were also some unexpected false positives as three participants responded that they recognized five images. Finally there were three more participants who failed to recognize the four pass-images. The pie chart signifies the percentages of participants who recognized all images correctly and "authenticated" (70%), those who did not recognize all images (15%) and finally false positives recognizing too many images (15%).

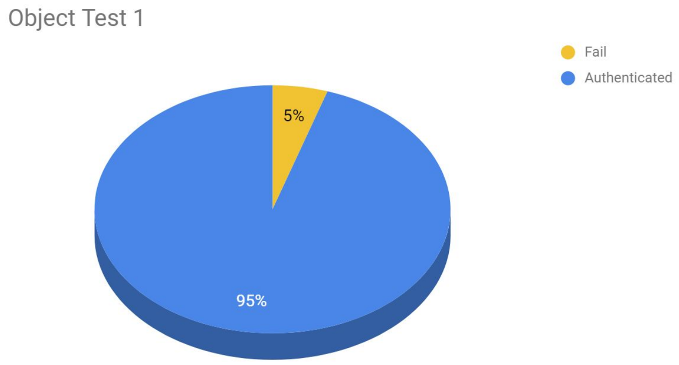
Abstract Test 2

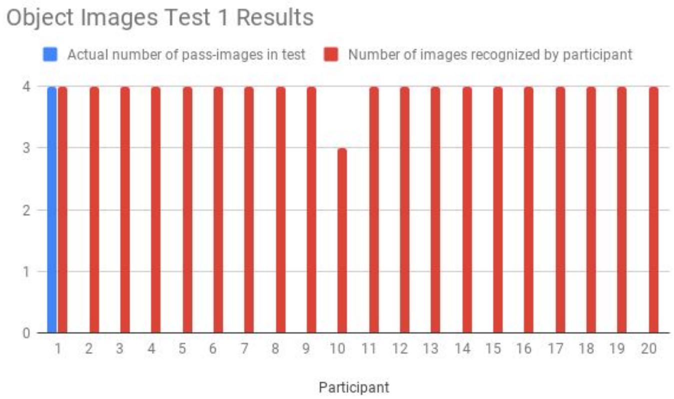




In Abstract test 2 we were looking for the rate of false positives. The chart on the right show each participant's response to the test and the chart on the right shows the percentages between the two outcomes. From the above charts it can be seen that seven participants claimed they recognized four images when there were only three. The other 13 participants responded with three recognized images matching what was expected. The pie chart shows that 35% of participant reports were false positives whereas 65% reported the displayed number of pass-images.

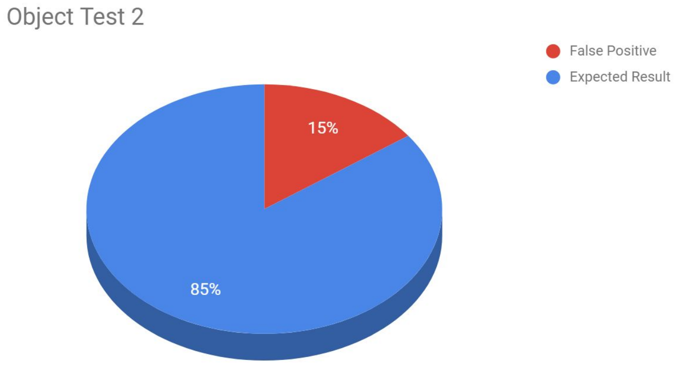
Object Test 1

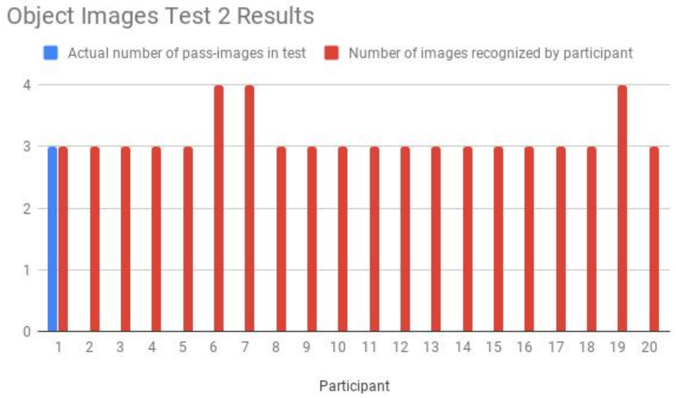




As can be seen by the bar chart all but one of the participants were able to recognize the four pass-images that were displayed. The pie chart on the left gives a success rate of 95% for determining the number of recognized images.

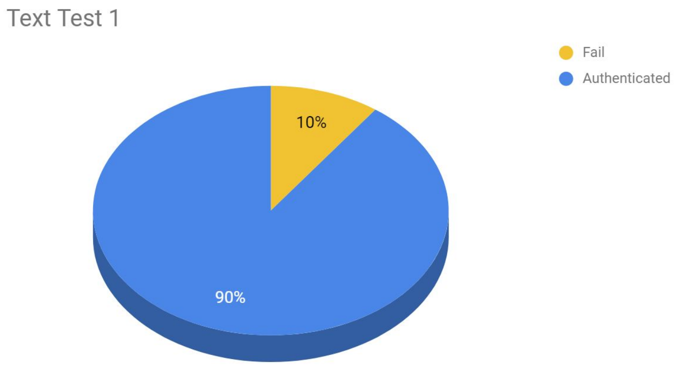
Object Test 2

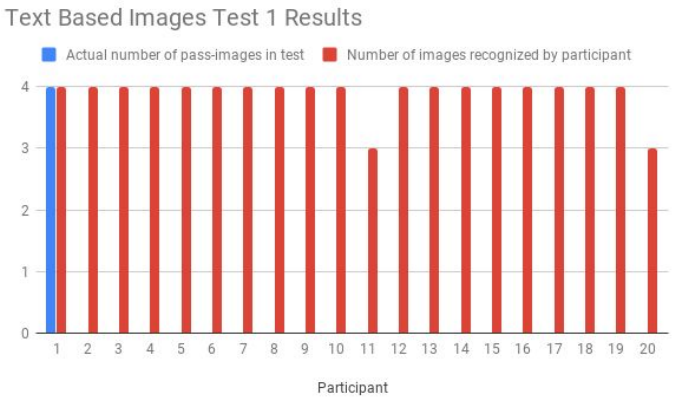




Object test 2 focused on false positives once again. The findings came out to be that three participants falsely identified an image as a pass-image, whereas the other 17 identified only the three that were there. Once again the pie chart gives percentages for the different outcomes. False positives came out to be 15% and 85% of users reported recognizing the three that were shown.

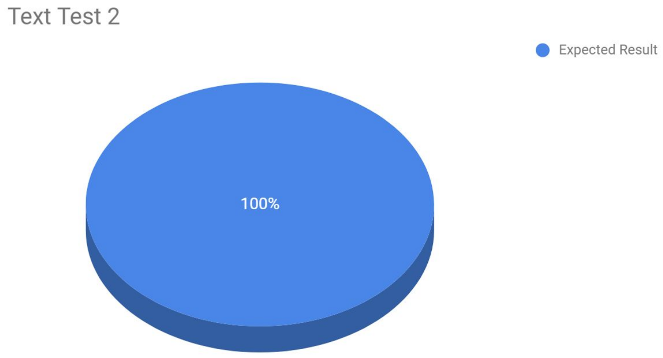
Text Test 1

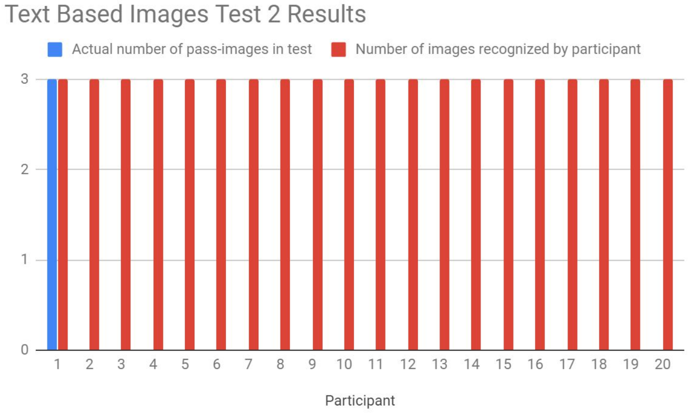




The text based test had 18 participants reporting that they had recognized all four pass-images. Two participants were only able to recognize 3 out of the four pass-images. This comes out to be a success rate of 90% to a fail rate of 10%.

Text Test 2





The second test involving test was to focus on false positives once again. As can be seen from the charts above there were zero false positives. This false positive rate of 0% may be attributed to the readability of the words shown. Most users had enough time to read the word on each image shown to them and could determine that way if it was one of their pass-images or not.

Qualitative Feedback

What Participants Liked Most

The following comments capture what the participants liked most:
"I liked the object images the most out of the three tests. I found that they were easier to recognize."
"I personally preferred the tests that had the dog, trees, etc (object images)."

What Participants Liked Least

The following comments capture what the participants liked the least:
"I did not like the abstract images test, those were definitely the hardest of the three."
"The text based tests were the worst."

Recommendations for Improvement

The following comments capture what the participants thought would improve the usability test:
"Use simpler object images. There is too much in the picture to look at."
"I think having personal photos mixed with random photos would be more secure."
"For improvements, why not try adding a test that uses random people's faces (or familiar friends). Use that as a key and then flash through random faces."

Recommendations

Based on the feedback received from participants and observations made by group members the following adjustments would take place. Future iterations of this system would contain different forms of pictures. For example, removing either the abstract images or the text based images from the image sets and opting for pictures of people's faces similar to Passfaces. It would be unlikely that users would be allowed to choose images of people that they know as that would reduce the security of the system. For instance any malicious actor who has the ability to learn about the user can find people they are close to and train themselves to recognize these people. It would however, be plausible if a General Adversarial Network (GAN) was used to create unique faces for users to learn and use for their pass-image. Using the GAN to create the faces would give users highly unique sets of images that no one else would have seen before. It would also be very useful when supplying the system with distractor images when actually attempting authentication. The benefits would mainly stem from the fact that the pool of images to pull from could be nearly infinite. The faces also having no personal connection to the user would make it more secure as malicious agents would have a difficult time pinpointing which faces were pass-images. Theoretically the GAN could also be used to generate not just images of faces that have never been seen before but also possibly landscapes or objects. This would then create an even greater pool of images that could be used in the proposed system.

Other recommendations included making the object images simpler and thus easier to be able to recognize. The idea behind this suggestion is logical as people can be distracted by what is in the image and may not recognize the image as one of theirs. This could also cause issue with images that look very similar. However in the ideal system being designed, the reaction to the image is a subconscious one that the user should not be aware of entirely. This would mean that it would not necessarily matter if the user consciously recognized the image as long as the P300 signal was captured.

Future Work

There are many ways we can further this study and make more informed design decisions that would give us more useful information about how to best design the system.

To continue this study, we should design and test a scenario to probe security of this system. While this BCI recognition-based system would, in theory, combat shoulder surfing, it would not eliminate it as the whole password could be pieced together after many exposures to a curious third-party actor. This could be done using a tester as a ‘spy’ who will observe someone’s login many times and try to piece together a successful password set that they can use to impersonate that individual.

Including a degradation test would help to demonstrate how repeatable our system is. We unfortunately lacked sufficient time to do so, but it would be as simple as retesting with the same password set and different distractor images about a week after the initial testing without re-exposure to the provided password set. This kind of test would likely show noticeable differences between the types of images -- abstract images that are hard to describe and can’t be associated with a single word would likely see a drop in success rates. This may also happen with the text images.

Including someone in the study group that is familiar with P300 use in research would be a benefit to the study. The aspect of the study that was most difficult to reason about and make decisions on was the use of P300, mostly due to our unfamiliarity with it.

Of course, in the long term, the study would benefit from being redone using some future BCI technology. The trials could be redesigned and redone with users using some commercial BCI headset/device that would allow a real-time feedback on P300 detection and help determine the usability and feasibility of a BCI recognition-based authentication scheme.

Finally, more tests could be conducted using more diverse image sets, such as running tests with faces, similar to PassFaces. We may see issues here with false positives as they are likely to show the bias towards faces that are attractive or of the same race. We could also expand to geographic locations, perhaps including some that hold some personal significance for the users, as that would help with both retention and recognition.

Limitations of Evaluation

There are a number of limitations to our evaluation of recognition based authentication with BCI technology that it is important to note. The most obvious limitation is that we did not have a brain computer interface to test. Instead of actually measuring the P300 wave we elected to test the users based off their self-reported recognition, meaning that when a user recognized an image they would take note and keep count. We attempted to alleviate this limitation to some degree by trying to see the number of false positives we could detect. This self-reporting mechanism is not necessarily representative of the results that may be shown by a true brain computer interface.

Flashing lights between the frequencies of 5 to 30Hz are most likely to trigger seizures [11]. Our test operated with a period of 350ms, or a frequency of 2.85Hz. Although this frequency was not within the 5 to 30Hz range we elected to not conduct the usability tests on any people with disabilities or illnesses such as epilepsy, brain damage, alzheimer's, or dementia in order to avoid any potential harm to our study participants. Similarly, as the study requires visual interaction, we cannot test with or apply this technology to individuals with severe visual impairments.

Our test banks of images consisted of sixteen to seventeen distractor images for each set of twenty images. The low number of pass-images that we had likely make our tested prototype susceptible to brute-force attacks as there was a small password space. The small password space also can contribute to weaknesses to shoulder-surfing, which is otherwise one of the key security concerns that our design addresses. Feedback we received hints that using a Generative Adversarial Network, or GAN, machine learning system may allow us to generate an infinite number of unique, never before-seen distractor images. This would likely allow us to sufficiently protect against these attacks, but we were not able to implement it for our study.

The actual creation of image sets is limited in a couple ways. Firstly, without considering the Generative Adversarial Network implementation for image generation, there is the issue of image distinctiveness. How do we guarantee that images are different enough from one another to elicit different responses where they should? Secondly, the system may be more effective if users have personally relevant images included in their password sets. Care would have to be taken to not include such images among the distractor images.

Finally, due to time considerations we were unable to measure out subjects' ability to recognize their pass-image set as time since training passed as a degradation study. In order to fully evaluate the usability of our proposed authentication system it may be necessary to conduct further research into this area.

References

[1] Li, Q. Ding, D. Conti, M. (2015). *Brain-Computer Interface applications: Security and privacy challenges.* Retrieved from https://ieeexplore.ieee.org/document/7346884/authors#authors

[2] A. B. Schwartz, X. T. Cui, D. Weber, D. W. Moran, "Brain-controlled interfaces: Movement restoration with neural prosthetics", *Neuron*, vol. 52, no. 1, pp. 205-220, 2006.

[3] Shih, J. Krusienski, D. Wolpaw, J. (March, 2012). *Brain-Computer Interfaces in Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497935/

[4] M. Yu, N. Kaongoen and S. Jo, "P300-BCI-based authentication system," 2016 4th International Winter Conference on Brain-Computer Interface (BCI), Yongpyong, 2016, pp. 1-4.

[5] Borkotoky, C., Galgate, S., & Nimbekar, S. B. (2008). Human computer interaction. Proceedings of the 1st Bangalore Annual Compute Conference on - Compute 08. doi:10.1145/1341771.1341797

[6] Emily Waltz, "From passwords to passthoughts, logging into your device with your mind," (August 2016), Retrieved from: https://spectrum.ieee.org/the-human-os/biomedical/devices/logging-into-your-devices-with-your-mind

[7] Ahn M, Lee M, Choi J, Jun SC. A review of brain-computer interface games and an opinion survey from researchers, developers and users. Sensors (Basel, Switzerland) 2014;14(8):14601–14633.

[8] Renaud, K. 2005a. Evaluating authentication mechanisms. In Security and Usability: Designing Secure Systems That People Can Use, L. Cranor and S. Garfinkel, Eds. O'Reilly Media, Stebastopol, C.A, Chapter 6, 103--128.

[9] Simon Thorpe. "Speed of processing in the human visual system," (1996), Retrieved from: https://www.nature.com/articles/381520a0

[10] Rik van Dinteren, Martijn Arns, Marijtje L. A. Jongsma, Roy P. C. Kessels. February 2014. "P300 Development across the Lifespan: A Systematic Review and Meta-Analysis". PLoS One. https://doi.org/10.1371/journal.pone.0087347.

[11] Giuseppe Erba. Shedding Light on Photosensitivity, One of Epilepsy's Most Complex Conditions. Retrieved April 18, 2019 from https://www.epilepsy.com/article/2014/3/shedding-light-photosensitivity-one-epilepsys-most-complex-conditions-0

[12] D. Davis, F. Monrose, and M.K. Reiter, “On User Choice in Graphical Password Schemes,” Proc. 13th Usenix Security Symp., Usenix Assoc., 2004, pp. 151–164. Retrieved from: https://users.ece.cmu.edu/~reiter/papers/2004/usenix2.pdf