# Passthoughts Authentication with Low Cost EarEEG

Max T. Curran[1], Jong-kai Yang[1], Nick Merrill[1] and John Chuang[1]
[1]BioSENSE Lab, School of Information, University of California, Berkeley
mtcurran@ischool.berkeley.edu, jong-kai.yang@berkeley.edu, ffff@berkeley.edu, chuang@ischool.berkeley.edu

***Abstract*—Personal and wearable computing are moving toward smaller and more seamless devices. We explore how this trend could be mirrored in an authentication scheme based on electroencephalography (EEG) signals collected from the ear. We evaluate this model using a low cost, single-channel, consumer grade device for data collection. Using data from 12 study participants who performed a set of 5 mental tasks, we achieve a 44% reduction in half total error rate (HTER) compared with a random classifier, corresponding to a 72% authentication accuracy in within-participants analyses and a 60% reduction and 80% accuracy in between-participant analyses. Given our results and those of previous research, we conclude that earEEG shows potential as a uniquely convenient authentication method as it is integrable into devices like earbud headphones already commonly worn in the ear, and the mental gestures generating the signal are invisible to would-be eavesdroppers.**

## I. Introduction

Personal and wearable computing devices are moving toward smaller sized or altogether absent display space and input methods aimed at simplified and effortless interactions. To maintain their utility, these devices allow access to personal and potentially sensitive information, therefore their security should also be at the forefront of necessary considerations. In this research, we explore earEEG as an authentication method with the potential to combine powerful protection, unobtrusive interaction, and a high degree of usability.

Secure methods for a system to authenticate a user can be grouped into three factors: inherence, something unique to the user like a fingerprint; knowledge, something only known to the user like a written password; and possession, a unique physical object owned by the user. One very commonly used method of authentication, a typed password, takes advantage of only the knowledge factor and by consequence a system can be easily tricked by an intruder who learns this password. Multi-factor methods that utilize multiple distinct factors, such as the practice of requiring both a memorized password and a code sent to a user's mobile phone via text message (knowledge and possession factors, respectively), greatly enhance security. As evidenced by this example however, multi-factor authentication can quickly become frustrating and cumbersome, adding additional unwanted steps to actions that may occur very frequently throughout a single day, such as logging in to an e-mail account or unlocking a mobile phone. Considering the proximity of wearables to the body and recent improvements in the collection and processing capabilities of biosignals, biosensory authentication methods can elegantly allow for multi-factor authentication without additional steps.

Electroencephalography (EEG), the measurement of neuronal electrical activity most typically non-invasively via electrodes (called channels) arranged on the scalp, is one such form of biosignal collection that can naturally employ both inherence and knowledge factors. EEG had its origins in clinical settings, used in the diagnosis and treatment of neurological disorders such as epilepsy, but has more recently moved into the general consumer market with several companies like Neurosky [1] and Emotiv [2] making wireless EEG devices available for personal use at prices as low as $100. One-step two-factor authentication using EEG signals has shown promise through a concept coined "passthoughts" which was first postulated, to our knowledge, in 2005 [3], then implemented and developed upon with expensive clinical-grade technologies [4], [5], [6] and later with consumer grade multi-channel [7] and single-channel [8] devices in 2011 and 2013, respectively. Our study most closely follows that of Chuang et al. (2013) who were able to achieve an authentication accuracy of approximately 99% using a single-channel (with a ground and reference ear clip) Neurosky Mindwave Mobile EEG headset. Their study involved participants performing a variety of mental tasks at a computer while wearing the headset. Some tasks were predefined and uniform across all participants like relaxed breathing, imaginary finger tapping, and an audio stimulus, while others included something secret and specific to each participant like imagining a chosen song, sport motion, or general thought. They concluded that single-channel EEG authentication is possible, despite lower a signal-to-noise (SNR) ratio, through customized task selections and signal similarity thresholds for each user and that task difficulty and level of engagement with the user should be leveraged in pursuit of usability for repeated use.

For the use of EEG signals in such applications as authentication to be feasible for widespread adoption and use, the stability, comfort, and appearance of the device must also be considered for the sake of the user and quality of data. The novel technique of measuring EEG signals via sensors placed only in or around the ear is one avenue by which these qualities may be achieved. The concept of earEEG has been explored and iterated upon in the last few years showing great promise achieving usable SNR ratios for the detection of auditory evoked potentials [9] and steady-state visual evoked potentials (SSVEP) in a brain-computer interface (BCI) reaching, on average, an accuracy level of approximately 88% [10]. While these studies utilized costly custom-fitted earpieces, generic silicon rubber earpieces may also be feasible [11] and wet electrodes may not be necessary to achieve desirable impedance levels [12]. Overall, earEEG is an exciting sensing modality because it affords the necessary

data quality to carry out brain sensing functionalities, and from the user's perspective because, compared with scalp-based EEG systems, it offers greater usability via discreet appearance and improved comfort. Already, many people regularly wear devices in their ears like earbud headphones, which could potentially integrate earEEG.

By marrying the two novel concepts of passthoughts authentication with single-channel earEEG data collection we intend to implement and examine a streamlined one-step multi-factor authentication process using a sensing device that is comfortable, discreet, cheap, and highly wearable. This study represents an initial step toward that goal as we investigate this potential using a minimally modified consumer grade device. We also discuss several ideas and opportunities for future development and refinement of this promising approach to effective and user-friendly security.

## II. Methods

### A. Study Procedures

12 UC Berkeley graduate student participants (7 male, 5 female) with a mean age of 28 $\pm$ 4.36 completed our study protocol approved by the UC Berkeley Committee for Protection of Human Subjects (CPHS). Study procedures began with an informed consent process, followed by a demographics questionnaire, a set up period with the earEEG device, completion of a set of 5 tasks presented on a laptop while EEG was recorded, and finally a post-experiment questionnaire. We used a Neurosky Mindwave Mobile wireless EEG headset, which is sold online to the general public for $99.99. Modifications made to the original included releasing the sensing electrode from the plastic forehead arm, removing the electrode, and replacing it by soldering a new 6 mm gold cup electrode onto the original wire. The gold cup electrode was bent to allow for a comfortable fit in a range of ear canal sizes. The device being worn is shown in Figure 1.

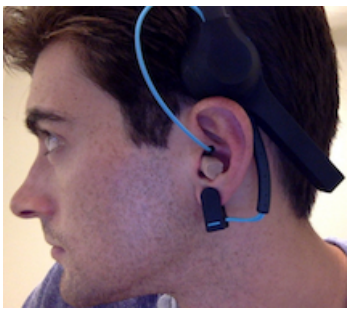

Fig. 1: Modified Neurosky Mindwave setup.

The setup process with a participant was fairly simple: the experimenter cleaned the electrode and the participant's ear canal with ethanol pads and cotton swabs, fastened the earlobe clip containing the ground and reference, applied a small amount of conductive gel to the electrode, and placed the sensing electrode in the ear canal against the superior wall (facing upward) with a rolled foam earplug placed beneath it to keep the electrode comfortably in place. The Mindwave device transmits data wirelessly via Bluetooth, so it was paired and the connection was confirmed before beginning the task phase of the experiment.

TABLE I: List of authentication tasks.

| Task Name | Description |
|---|---|
| Breath | Relaxed breathing with eyes closed |
| Song | Imagine a chosen song with eyes closed |
| Listen | Listen to a 40 Hz tone with eyes closed |
| Face | Imagine a chosen face with eyes closed |
| Cube | Imagine a displayed cube is rotating with eyes open |

Table I lists the tasks performed by participants. Instructions for tasks were presented visually using PsychoPy [13] and read aloud verbally by the experimenter. Each task was recorded during two sets of 5 trials each to lessen boredom effects and each trial was 12 seconds in length. Our paradigm synced the earEEG recording with the start of each trial. We primarily chose tasks that showed promise in previous passthoughts experiments, though generally we found little research to suggest which mental tasks result in the most robust EEG signals. We attempted to include a variety of tasks to hopefully draw on different EEG signals. The breath, song, listen, and face tasks were performed with the participants' eyes closed, while the cube task was performed with the participants' eyes open. In an authentication situation, the knowledge of which task to perform could be considered the supplemental knowledge security factor. In order to explore this further though, two of the tasks, the song and face tasks, involved an additional unique choice by the participant. We asked participants to sit in a comfortable position and remain as still as possible for all tasks with the intention of minimizing signal pollution by electromyographic signals (EMG) (signals generated by muscle movement). Participants used a wireless remote held comfortably in their laps to begin each task when they were ready.

### B. Authentication Analysis

We performed two main analyses to test authentication: a within-participants analysis to see how well authentication would fare using different tasks of only a single user, and a between-participants analysis to test authentication using data from all users' tasks. In both analyses, we assessed the authentication ability of each task for each participant by calculating the measures of false rejection rate (FRR), the rate of rejecting the correct user, false acceptance rate (FAR), the rate of accepting the incorrect user, and half total error rate (HTER), the average of the FRR and FAR.

We analyzed the EEG signals collected during the tasks using a support vector machine (SVM) classifier. Since past work has shown that classification tasks in EEG-based BCI are linear [14], we used LIBLINEAR, [15], a popular linear SVC kernel. For each task, for each participant, 120 seconds of data was collected in total across 10 trials of 12 seconds each. We initially tried analyzing all 12 seconds of data per trial,

but found that removing the first 2 seconds and last second of each trial to account for the transition to and from performing a given task improved our results. Following [16], we used logarithmic binning, an approach known to produce small feature vectors with good linear classifiability. This approach allowed us to average multiple power spectra over time. We selected a feature resolution of 3 seconds, as this was the shortest duration that produced good results in our test, and then applied the logarithmic binning. After this preprocessing we had 30 samples per participant, per task.

For the within-participants analysis, for each participant for each task, 15 samples were randomly selected as training set A, and 15 samples as testing set A. From the other 120 samples (30 samples x 4 other tasks), 15 were randomly selected as training set B, and another 15 as testing set B. The SVM was then trained on classifying between training set A as a model of correct authentication, and training set B as a model of correct rejection. The FRR was calculated by testing the SVM on testing set A (samples from the same task), resulting in a list of 15 0's (authenticate) and 1's (reject) and taking the mean of this list. Similarly, the FAR was calculated by testing the trained SVM on testing set B (samples from other tasks) and taking 1 minus the mean of the resulting list of 0's and 1's. We ran this random sampling and testing 1,000 times for each participant/task.

The process was very similar for the between-participants analysis, differing only in that the 15 samples used for training set B and testing set B were randomly selected from all other participants' samples rather than from within the same participant's samples.

Finally, the HTER was calculated by averaging the FRR and FAR for each task/participant pair. As these are binary classifiers we can evaluate our results as falling somewhere between a completely random binary classifier, which would have an HTER of 0.50 and authentication accuracy of 50%, and a perfect classifier, which would have an HTER of 0 and authentication accuracy of 100%.

## III. Results

Results following the within-participants analysis method discussed above are shown in Table II. These results are presented by task, averaged across all participants with standard errors shown. The Best Task HTER measure was calculated as the mean across participants using only each participant's lowest HTER among their tasks. The breakdown of best tasks selected for this measure was: the breath task for 3 participants, the listen task for 3 participants, the face task for 3 participants, and the cube task for 3 participants. The song task did not perform best for any participant.

Results following the between-participants analysis method are shown in Table III. These results are again presented by task, averaged across all participants with standard errors shown and the Best Task HTER measure was calculated in the same way. The breakdown of best tasks selected for this measure in this analysis was: the breath task for 2 participants, the listen task for 5 participants, the face task for 3 participants, and the cube task for 2 participants. Again, the song task did not elicit the best HTER for any participants, though was a close second for a few.

TABLE II: Within-participant authentication results, means across participants with standard errors.

| Task | FRR | FAR | HTER |
|---|---|---|---|
| Breath | 0.380 ± 0.024 | 0.386 ± 0.022 | 0.383 ± 0.023 |
| Song | 0.425 ± 0.021 | 0.434 ± 0.017 | 0.430 ± 0.019 |
| Listen | 0.360 ± 0.035 | 0.385 ± 0.034 | 0.373 ± 0.034 |
| Face | 0.410 ± 0.031 | 0.416 ± 0.032 | 0.413 ± 0.031 |
| Cube | 0.367 ± 0.030 | 0.399 ± 0.023 | 0.383 ± 0.026 |
| **Best Task HTER** | **0.281 ± 0.030** | | |

TABLE III: Between-participant authentication results, means across participants with standard errors.

| Task | FRR | FAR | HTER |
|---|---|---|---|
| Breath | 0.278 ± 0.031 | 0.323 ± 0.037 | 0.300 ± 0.033 |
| Song | 0.288 ± 0.025 | 0.330 ± 0.026 | 0.309 ± 0.025 |
| Listen | 0.199 ± 0.034 | 0.258 ± 0.034 | 0.228 ± 0.034 |
| Face | 0.238 ± 0.027 | 0.290 ± 0.030 | 0.264 ± 0.028 |
| Cube | 0.237 ± 0.028 | 0.295 ± 0.031 | 0.266 ± 0.029 |
| **Best Task HTER** | **0.200 ± 0.03** | | |

The post-experiment questionnaire results are displayed in Figure 2. Each question answered via a a 7-point likert-type scale. The questions asked were: "Please rate each of the experimental tasks on ease of performing." (1 very difficult, 7 very easy); "Please rate each of the experimental tasks on how engaging/interesting they were to perform." (1 very boring, 7 very engaging); "Please rate each of the experimental tasks on how easy they would be to repeat often." (1 very difficult to repeat, 7 very easy to repeat); and "Please rate each of the experimental tasks on how likely you would be to use each in a real-world authentication setting." (1 not at all likely to use, 7 very likely to use).

## IV. Discussion of Results

In the within-participants analysis, the listen task performed best with an HTER of 0.373 followed by the breath and cube tasks tied at 0.383. Using the best performing task for each participant resulted in a much-improved HTER of 0.281. This is a 43.8% reduction in error rate compared to a random classifier, corresponding to a 71.9% authentication accuracy.

In the between-participants analysis, the listen and cube tasks performed best with HTERs of 0.228 and 0.264, respectively. Selecting the best task per participant again improved this value to 0.200, a 60% reduction in error rate compared to a random classifier, corresponding to an 80% authentication accuracy.

Overall, the within-participants authentication performed worse than the between-participants analogue. This result was expected as it replicates previous scalp-based passthoughts work by Chuang et al. that suggested single-channel EEG is better at distinguishing between users compared to distinguishing between signals of a single user. These performances were worse than previous scalp-based experiments however, likely

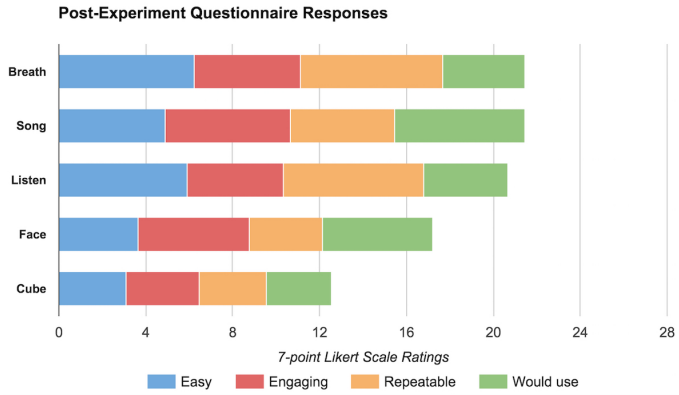


Fig. 2: Post-experiment questionnaire mean responses on a 7-point likert-type scale.

due to the increased distance and absorption material between the electrode and the cortex that produces the EEG signals.

The post-experiment questionnaire results indicated that participants generally felt the tasks were easy to perform, engaging, repeatable, and were at least somewhat likely to use them for authentication. Two of the best performing tasks in authentication, breath and listen, scored highly in ease of use and repeatability and moderately in engagement and likeliness of use. The other best performing task was the cube task, though it scored the lowest in all four of these domains. The song task scored highly in all four domains, but performed the worst for authentication. It is important to note that in order to achieve the Best Task HTERs users would not be able to choose which task to perform as the best performing task would be selected for each individually. In our results there was not one single task that was best across participants, at most 5 out of 12 of the participants had the listen task selected for them in the between-participants analysis.

There are many avenues for continued research on this topic with potential for improvement and expansion of the results. Firstly, our data was collected using a fairly rudimentary setup and the signal quality could be improved with custom-fit earpieces or by adding additional sensors within the ear canal or on the outside of the ear. Also, the 5 tasks we chose to test here may not be the best for producing unique, distinguishable signals. Other tasks may result in lower HTERs and higher authentication accuracies; exploration of a much larger range of different tasks to use would be very useful.

A major limitation to our study is that it does not address how these results scale with number of users. Not only would the distinguishability between users change as more users are added into the system, but a given user's best task may change as well. One possibility to address this potential disadvantage would be to use combinations of tasks, performed in a specified sequence for each user. This improvement would likely come at a cost to usability however, as depending on what situations this authentication method is used in users generally want to authenticate into their devices with as little time and effort as possible.

## V. Conclusions

Using a consumer grade, single-channel EEG device, minimally modified to sense from the ear, we analyzed the efficacy of a simple scheme for user authentication. Specifically, we achieved 43.8% reduction in half total error rate for within-participants authentication, and a 60% reduction for between-participants, compared with a random classifier. While the achieved rates are not immediately viable for real-world authentication applications, the results do show this method has promise. We present a few avenues future work might pursue in improving authentication for earEEG devices.

*This research was supported in part by a Google Faculty Research Award, the Hewlett Foundation through the UC Berkeley Center for Long-Term Cybersecurity, and the National Science Foundation under award CCF-0424422 (TRUST).*

## References

[1] [Online]. Available: http://www.neurosky.com
[2] [Online]. Available: http://www.emotiv.com
[3] J. Thorpe, P. C. van Oorschot, and A. Somayaji, "Pass-thoughts: authenticating with our minds," in *Proceedings of the 2005 workshop on New security paradigms*. ACM, 2005, pp. 45–56.
[4] M. Poulos, M. Rangoussi, N. Alexandris, A. Evangelou *et al.*, "Person identification from the eeg using nonlinear signal classification," *Methods of information in Medicine*, vol. 41, no. 1, pp. 64–75, 2002.
[5] S. Marcel and J. R. Del Millan, "Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 743–752, 2007.
[6] R. Palaniappan, "Two-stage biometric authentication method using thought activity brain waves," *International Journal of Neural Systems*, vol. 18, no. 01, pp. 59–66, 2008.
[7] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (eeg) based authentication," in *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*. IEEE, 2011, pp. 442–445.
[8] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore i am: Usability and security of authentication using brainwaves," in *Financial Cryptography and Data Security*. Springer, 2013, pp. 1–16.
[9] P. Kidmose, D. Looney, and D. P. Mandic, "Auditory evoked responses from ear-eeg recordings," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 586–589.
[10] Y.-T. Wang, M. Nakanishi, S. L. Kappel, P. Kidmose, D. P. Mandic, Y. Wang, C.-K. Cheng, and T.-P. Jung, "Developing an online steady-state visual evoked potential-based brain-computer interface system using eareeg," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 2271–2274.
[11] P. Kidmose, D. Looney, L. Jochumsen, and D. P. Mandic, "Ear-eeg from generic earpieces: A feasibility study," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 543–546.
[12] S. L. Kappel and P. Kidmose, "Study of impedance spectra for dry and wet eareeg electrodes," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 3161–3164.
[13] J. W. Peirce, "Psychopy-psychophysics software in python," *Journal of neuroscience methods*, vol. 162, no. 1, pp. 8–13, 2007.
[14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, p. R1, 2007.
[15] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
[16] N. Merrill, T. Maillart, B. Johnson, and J. Chuang, "Improving physiological signal classification using logarithmic quantization and a progressive calibration technique," 2015.