Exploring the Feasability of Multifactor Authentication with EarEEG

1st Author, 2nd Author
Affiliation
1st e-mail, 2nd e-mail

3rd Author
Affiliation
e-mail

4th Author
Affiliation
e-mail

ABSTRACT

Multifactor authentication presents a robust method to secure our private information, but typically requires multiple actions by the user resulting in a high cost to usability and limiting adoption. A usable system should also be unobtrusive and inconspicuous. We present and discuss a system with the potential to engage all three factors of authentication (inherence, knowledge, and possession) in a single step using an earpiece that implements brain-based authentication using electroencephalography (EEG). We demonstrate its potential by collecting EEG data using manufactured custom-fit earpieces with embedded electrodes and testing a variety of authentication scenarios. Across all participants best-performing passthoughts, we are able to achieve 0% false acceptance and 0.36% false rejection rates, an overall accuracy of 99.82%, using one earpiece with three electrodes. Furthermore, we find no successful attempts simulating impersonation attacks. We also report perspectives from our participants and discuss directions for future work to explore earEEG for the masses. Our results indicate that a relatively inexpensive system using a single electrode-laden earpiece could provide a discreet, convenient, and robust method for one-step, three-factor authentication.

Author Keywords

usable security; multi-factor authentication; wearable authentication; passthoughts; biosensing

INTRODUCTION

It is well appreciated by experts and end-users alike that strong authentication is critical to cybersecurity and privacy, now and into the future. Unfortunately, news reports of celebrity account hackings serve as regular reminders that the currently dominant method of authentication in consumer applications, single-factor authentication using passwords or other user-chosen secrets, faces many challenges. Major industry players such as Google and Facebook have strongly encouraged their users to adopt two-factor authentication (2FA). However, submitting two different authenticators in two separate steps has frustrated wide adoption due to its additional hassle to users. The Apple iPhone, for instance, already supports device unlock using either a user-selected passcode or a fingerprint. The device could very well support a two-step two-factor authentication scheme if desired. However, it is easy to understand why users would balk at having to enter a passcode *and* provide a fingerprint each time they want to unlock their phone.

"One-step two-factor authentication" has been proposed as a new approach to authentication that can provide the security benefits of two-factor authentication without incurring the hassle cost of two-step verification. In this work we undertake, to the best of our knowledge, the first ever study and implementation of one-step, *three*-factor authentication. In computer security, authenticators are classified into three types: knowledge factors (e.g., passwords and PINs), possession factors (e.g., physical tokens, ATM cards), and inherence factors (e.g., fingerprints and other biometrics). By taking advantage of a physical token in the form of personalized earpieces, the uniqueness of an individual's brainwaves, and a choice of mental task to use as one's "passthought", we seek to achieve all three factors of authentication in a single step by the user.

Furthermore, the form factor of an earpiece combats the flaw of the conspicuous and obtrusive nature of traditional EEG systems worn on the scalp. Technology worn in the ear is already a socially accepted practice in many cultures, with examples like earphones or bluetooth headsets.

We find that we can achieve very low false acceptance rates (FAR) and false rejection rates (FRR) with a single, three-electrode earpiece. Interestingly, we found that performance improves in the ear compared to a single electrode on forehead above the left eye, a site typically used for collecting EEG data. We also find that passthoughts could not be spoofed by imposters, even when the imposters had the user's earpiece and knew their chosen passthoughts. These results improve upon previous work on earEEG-based passthoughts systems by testing custom-made earpieces with multiple embedded electrodes which achieve above 99% performance, an approximately 20% improvement [8]. We also detail important aspects of the system including measurements of the electrodes' electrical impedances and the results of various classifier training and testing schema to probe each authenetication factor. Finally we discuss interesting patterns in our promising results, and future work to investigate the robustness of passthoughts in day-to-day life.

RELATED WORK

Passthoughts and Behavioral Authentication
The use of EEG as a biometric signal for user authentication has a relatively short history. In 2005, Thorpe et al. motivated and outlined the design of a passthoughts system [28]. Since 2002, a number of independent groups have achieved 99-100% authentication accuracy for small populations using research-grade and consumer-grade scalp-based EEG systems [23, 17, 3, 7].

The concept of in-ear EEG was introduced in 2011 with a demonstration of the feasibility of recording brainwave signals from within the ear canal [14]. The in-ear placement can produce signal-to-noise ratios comparable to those from conventional EEG electrode placements, is robust to common sources of artifacts, and can be used in a brain-computer interface (BCI) system based on auditory and visual evoked potentials [13]. One study attempted to demonstrate user authentication using in-ear EEG, but was only able to attain an accuracy level of 80%, limited by their use of a consumer-grade device with a generic fit, single-electrode earpiece [8]. Our study improves on these results by incorporating a physical token (a possession factor) in the form of custom-fitted earpieces with three electrodes each vs. the single channel used in previous work. Furthemore, we use the OpenBCI system which allows us to capture raw data (the device used in previous work used an unknown proprietary algorithm), simultaneously record from the scalp for validation, and report impedance measures important to data quality and authentiation assessment.

Behavioral authentication methods such as keystroke dynamics and speaker authentication can be categorized as one-step two-factor authentication schemes. In both cases, the knowledge factor (password or passphrase) and inherence factor (typing rhythm or speaker's voice) are employed [20]. In contrast, the Nymi band supports one-step two-factor authentication via the inherence factor (cardiac rhythm that is supposed to be unique to each individual) and the possession factor (the wearing of the band on the wrist) [21]. However, as far as we know, no one has proposed or demonstrated a one-step three-factor authentication scheme.

Usable Authentication
When proposing or evaluating authentication paradigms robustness against imposters is often the first consideration, but the usability of these systems is of equal importance as they must conform to a person's needs and lifestyle to warrant adoption and prolonged use. The authors of [26] describe usability issues with knowledge-based systems like alphanumeric passwords, in particular that it should not be the burden of the user to remember complex passwords that have to be frequently changed. In [4], researchers analyzed some of the complexities applying interface human factors heuristics to authentication, and indicate the importance of social acceptibility, learnability, and simplicity of authentication methods. Technology worn on the head faces particular usability issues; in their analysis of user perceptions of such devices, the authors of [11] identified design, usability, ease of use, and obtrusiveness among the top 10 concerns of users' online postings, and qualitatively comments around comfort and "looking weird".

In the proposed system we seek to incorporate the perspectives of this research. Passthoughts such as relaxed breathing or imagining favorite song are easy for a user to remember and perform, and the earpiece form factor of our system is intended to address wearability and social perception concerns.

Mobile and wearable technologies' continuous proximity to the user's body provides favorable conditions for unobtrusively capturing biometrics for authentication. Many such uses have been proposed that embrace usability like touch-based interactions [27, 12] and walking patterns [15] using mobile phones, as well as identification via head movements and blinking in head-worn devices [24]. However, these typically draw only from the inherence factor. [6] proposed an inherence and knowledge two-factor method for multi-touch mobile devices based on a user's unique finger tapping of a song, though it may be vulnerable to "shoulder surfing": imposters observing and mimicking the behavior. By comparison, passthoughts are largely invisible to attackers. Furthermore, this system also presents a possible solution to "rubber-hose attacks" (forceful coercion to divulge a password): a thought, which is secret (and thus changeable), but has a particular expression unique to an individual, the specific performance of which cannot be described (and thus cannot be coerced).

METHODS

Study Overview
Seven male, right-handed participants (P1-P7), five students and two non-students, completed our study protocol approved by our local ethics review board. Though this sample is relatively homogenous and greater diversity is necessary for a larger real-world feasibility assessment, this quality interestingly functions to strengthen the results of this initial exploration (see details in Discussion). After participants' 3D ear molds were obtained and their fit and electrical impedances were checked, the collection of data for authentication analyses commenced. Data collection consisted of participants completing a demographics questionnaire, a set up period with the OpenBCI system and earpieces with a second impedance check, their performance of nine mental tasks, and finally a post-experiment questionnaire.

Earpiece Design and Manufacturing
To produce custom ear impressions we first cleaned subjects' ears, placed a cotton ball with a string attached into the ear canal, and injected silicon into the canals. When the silicon dried after a few minutes, the string was pulled to remove the impression from the ear canal. This impression was then scanned with a 3D scanner and the resulting scan modified based on some heuristics to achieve a comfortable fit and to ensure the intended electrode sites will make good contact with the skin. Channels were created in the 3D model to allow wire leads and associated EEG electrodes as well as a plastic tube to deliver audio. This 3D model was then sent to a 3D printer after which wires, leads, and associated AgCl electrodes were installed. The positions of the earpiece
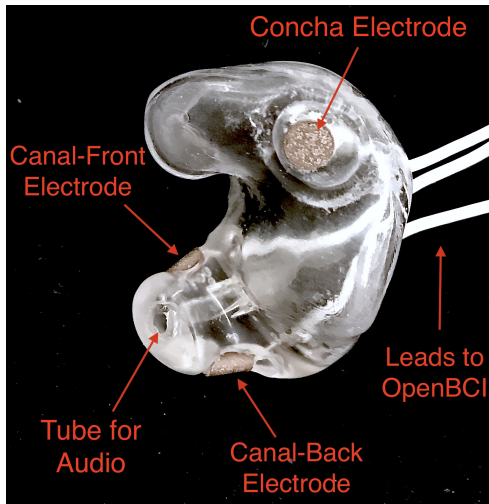
Figure 1. Labeled photo of one of our manufactured custom-fit earpieces with thre embedded electrodes located in the concha, front-facing (anterior) in the ear canal, and back-facing (posterior) in the ear canal.

electrodes were simplified from those described in [19]. We reduced the number of canal electrodes in order to prevent electrical bridging and positioned them approximately 180 degrees apart in the canal (posterior/back and anterior/front locations in the canal). One other electrode was placed in the concha. An example of one of the manufactured earpieces is shown in Figure 1.

Mental Tasks

We selected a set of mental tasks based on findings in related work regarding the relative strengths of different tasks in authentication accuracy and usability as reported by participants. [7, 8] Furthermore, given the in-ear placement of the electrodes and therefore the proximity to the temporal lobes containing the auditory cortex, we tested several novel authentication tasks based specifically on aural imagery or stimuli. The nine authentication tasks and their attributes are listed in Table 1. Our strategy was to select tasks that captured a diversity across dimensions of external stimuli, involving a personal secret, eyes open or closed (due to known effects on EEG), and different types of mental imagery.

Data Collection Protocol

All sites were cleaned with ethanol prior to electrode placement and a small amount of conductive gel was used on each electrode. For EEG recording we used an 8-channel Open-BCI system [18] which is open-source and costs about 600 USD; an alternative to medical-grade EEG systems (which cost >20,000 USD), with demonstrated effectiveness [9]. The ground was placed at the center of the forehead, approximately AFz according to the 10-20 International Standard for Electrode Placement (ISEP), and reference on the left mastoid. The AFz ground location was intentional to not bias left or right ear recordings, though future systems using one ear only should test relocating the ground to a site on one ear (e.g. the earlobe). Six channels were used for the three electrodes on each earpiece (shown in Figure 1); for the remaining two channels one AgCl ring electrode was placed

on the right mastoid for later re-referencing, and one at approximately Fp1 (ISEP location above the left eye) to validate the data collected in the ears against a common scalp-based placement. Before beginning the experiment, the data from each channel was visually inspected using the OpenBCI interface. Audio stimuli were delivered through small tubes in the earpieces.

During the experiment, participants were seated in a comfortable position in a quiet room facing a laptop on which the instructions and stimuli were presented using PsychoPy [22]. All tasks were performed for five trials each, followed by another set of five trials each to reduce boredom and repetition effects. Each trial was 10 seconds in length, for a total of 10 trials or 100 seconds of data collected per task. The instructions were read aloud to participants by the experimenter, and participants advanced using a pointer held in their lap to minimize motion artifacts in the data. The experimenter also recorded the participant's chosen secrets for the *sport*, *song*, *face*, *speech*, and *sequence* tasks and reminded the participant of these for the second set of trials. After EEG data collection, participants completed a usability questionnaire.

ANALYSIS

Data Validation

We confirm the custom-fit earpieces are able to collect quality EEG data via two metrics: low impedances measured for the ear electrodes, and alpha-band EEG activity attenuation when a participant's eyes were open versus closed.

It is important that the electrical impedances achieved for electrodes are low (<10 kOhm) to obtain quality EEG signals. Table 2 below summarizes the impedances across the seven participants' six ear channels. With the exception of a few channels in select participants, impedances achieved were good overall. Most of the recorded impedances of the earpiece electrodes were less than 5 kΩ, a benchmark used widely in previous ear EEG work, and all except two were less than 10 kΩ. Nonetheless, the data from all electrodes were tested in our other data quality test.

For the alpha-attenuation test, data from the *breathe* task was compared with that of the *breathe - open* task. It is a well-known feature of EEG data that activity in the alpha-band (approx. 8-12 Hz) increases when the eyes are closed compared to when the eyes are open. This attenuation is clearly visible even in just a single trial's data from our earpieces and matches that seen in our Fp1 scalp electrode data. Figure 2 shows the alpha attenuation in the left ear channels, as well as Fp1 of one participant as an example. We see the same validation in the right ear channels.

Classification

Since past work has shown that classification tasks in EEG-based brain-computer interfaces (BCI) are linear [10], we used XGBoost, a popular tool for logistic linear classification [5], to analyze the mental task EEG data. Compared to other linear classifiers, XGBoost uses gradient boosting in which an algorithm generates aa decision tree of weak linear classifiers that minimizes a given loss function. Gradient boosting

Task	Description	Stimuli?	Secret?	Eyes	Imagery
Breathe	Relaxed breathing with eyes closed	No	No	Closed	None
Breathe - Open	Relaxed breathing with eyes open	No	No	Open	None
Sport	Imagine attempting a chosen physical activity	No	Yes	Closed	Motor
Song	Imagine hearing a song	No	Yes	Closed	Aural
Song - Open	Song task, with eyes open	No	Yes	Open	Aural
Speech	Imagine a chosen spoken phrase	No	Yes	Closed	Aural
Listen	Listen to noise modulated at 40 Hz	Yes	No	Closed	None
Face	Imagine a chosen person's face	No	Yes	Closed	Visual
Sequence	Imagine a chosen face, number, and word on timed cues	Yes	Yes	Open	Visual

Table 1. The nine authentication tasks and their properties. We selected tasks with a variety of different properties, but preferred tasks that did not require external stimuli, as the need to present such stimuli at authentication time could present challenges for usability and user security.

Impedances [kΩ]

	Left ear			Right ear		
P	**C**	**F**	**B**	**C**	**F**	**B**
1	4	4	4	<1	4	3
2	9	5	4	3	4	4
3	4	5	4	9	6	9
4	4	5	4	3	16	9
5	9	20	7	3	7	9
6	5	8	2	1	1	9
7	2	9	8	7	5	6

Table 2. Electrical impedances measured for concha (C), front (F) and back (B) earpiece electrodes.

generally improves linear classification results without manually tuning hyperparameters.

To produce feature vectors, we took slices of 100 raw values from each electrode (about 500ms of data), and performed a fourier transform to produce power spectra for each electrode during that slice. We concatenated all electrode power spectra together. No dimensionality reduction was applied. For each task, for each participant, 100 seconds of data were collected in total across 10 trials of 10 seconds each, resulting in 200 samples per participant, per task.

We trained the classifier such that positive examples were from the target participant and target task, and negative examples were randomly selected tasks from any other participant. From this corpus of positive and negative samples, we withheld one third of data for validation. The remaining train-set was used to cross-validate an algorithm over 100 rounds on different splits of the data. The results of each cross-validation (CV) steps were used to iteratively tweak classifier parameters.

For the predictions, the evaluation regards the instances with prediction value larger than 0.5 as positive instances, and the others as negative instances. After updating classifier parameters, the classifier was tested on the withheld validation set. Since negative examples far outweigh positive examples in this dataset, XGBoost atuomatically optimized using the error hyperparameter. Over a set of E examples containing E_W wrong examples $E_W \subset E$, XGBoost's binary classification error rate ϵ is calculated as
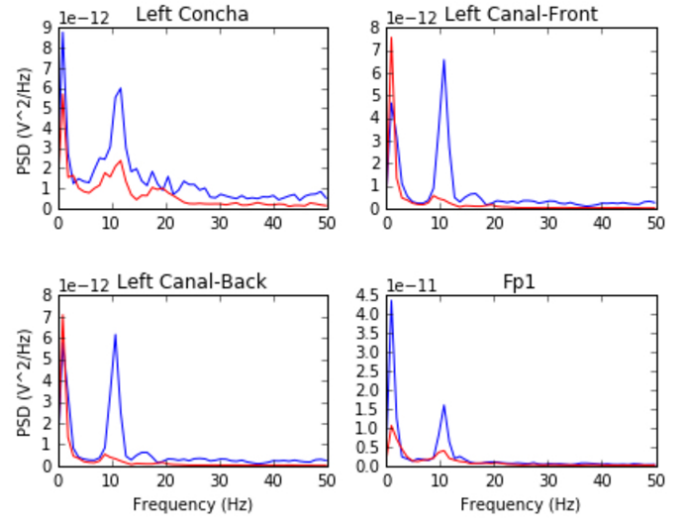
$$\epsilon = E_W/E$$

Figure 2. Alpha-attenuation (8-12 Hz range) in left ear and Fp1 channels, referenced at left mastoid. Red indicates breathing data with eyes open, blue indicates the same task with eyes closed.

We calculated FAR and FRR from these results. Over false attempts FA of which some subset FA_S were successful, and true attempts TA over which some subset TA_U were unsuccessful:

$$FAR = FA_S/FA$$

$$FRR = TA_U/TA$$

To further test the robustness of the system, we also conducted a leave one out process for the best performing tasks in which each participants FAR was calculated once with each other participant left out (e.g. CV for P1 with P2 left out, then CV for P1 with P3 left out, etc. for every participant combination).

RESULTS

Electrode Configuration
For each configuration of electrodes, we calculated the mean FAR and FRR across all participants using each task as the passthought (Figure 3). Incorporating all electrodes data resulted in the lowest FAR, followed by the combined right and
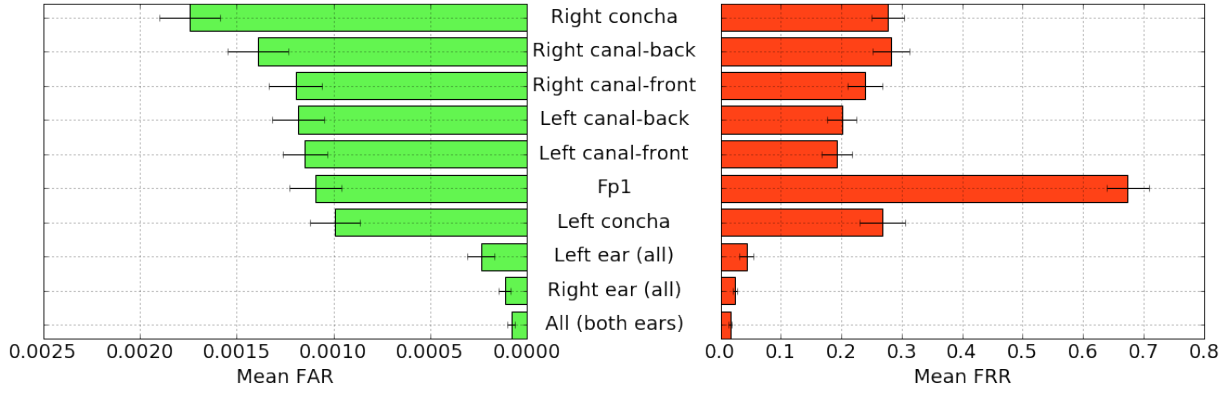
Figure 3. Mean FAR and FRR by electrode configuration across all participants and tasks. All electrodes (Fp1, right, and left ear channels) combined achieved the best FAR score, followed by the right and left ear electrodes combined, respectively.

left ear electrodes, respectively. For left ear, right ear, and all electrode configurations every participant had at least one task with 0 FAR/FRR. Counter to our expectations, Fp1 does not perform as well as most ear electrodes, though overall these reported FAR rates are $<<1\%$.

For each position, FAR was about ten times lower than FRR, which is preferable for authentication, as false authentications are more costly than false rejections.

Our results indicate acceptable accuracy using data from the left ear alone. This corresponds to a desirable scenario, in which the device could be worn as a single earbud. As such, we focus on results from only the left ear in the following analyses.

Authentication Results

Using only data from the left ear electrodes, the FARs and FRRs of each task for each participant are shown in Tables 3 and 4, respectively. We find at least one task for each participant that achieves 0% FAR, and for five participants a task where both the FAR and FRR are 0%. Each task achieved perfect 0 FAR and FRR for at least one participant, notably *breathe* and *song - open* achieved perfect FAR and FRR for three out of seven participants.

FAR and FRR results by task are shown in Figure 4, averaged across participants. Across all tasks, the sport task produced the lowest FAR. Specifically, it produced 0.0% FAR for all seven participants, with a corresponding 1.8% FRR. This suggests that the authentication scheme can work very well even if we limit the passthoughts to just a single task category, where the users could choose a personalized secret for that task. Interestingly, tasks like *breathe* and *breathe - open* performed very well despite lacking a personalized secret, indicating that even when the task may be the same across participants our classifier was still able to distinguish between them.

As an omnibus metric, the half total error rate (HTER) is defined as the average of the FAR and FRR:

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

and from this we estimate authentication accuracy, ACC, as:

$$ACC = 100 * (1 - HTER)$$

Using our best performing tasks' FARs, averaging 0% and these tasks' associated FRRs, averaging 0.36%, we estimate a best-case authentication accuracy of 99.82%.

These results thus far establish good performance in our default training strategy, in which we count as negative examples recordings from the wrong participant performing any task. For comparison, we report three other analyses with differing negative examples which serve to isolate and test the inherence and knowledge factors: the correct task recorded from the wrong participant (relies on inherence only), the wrong task recorded from the correct participant (relies on knowledge only), and a combination of these two. Positive examples were always the correct participant performing the correct task.

Overall, our default training strategy which engages both knowledge and inherence factors achieves the lowest FAR (Table 5). The FAR in the inherence-only scenario is ten times higher, and the knowledge-only FAR is one hundred times higher, though for all scenarios FAR is less than 1%. However, FRR is *lower* in the inherence-only training strategy than the default. The highest FRR is in the combined negative example case, though FAR remains low.

Our "leave one out" analysis with participants' best tasks maintained 0 FAR across all participant combinations.

Usability

Before the end of othe session, participants completed a usability questionnaire. This questionnaire asked them to rate each mental task on four 7-point likert-type scales: ease of use, level of engagement, perceived repeatability, and likeliness to use in a real-world authentication setting. The *breathe* and *listen* tasks were rated easiest to use (μ=6.75), *Sequence*, *song*, and *song - open* as most engaging (μ=5), *Breathe* (μ=7) and *listen* (μ=6.75) as most repeatable. For likeliness to use for authentication, *song - open* (μ=5) and *sequence* (μ=4.25) were rated highest, though modestly. Participants also ranked

Task	P1	P2	P3	P4	P5	P6	P7
Breathe	0	0	0	0	0.0002	0.0004	0
Breathe - open	0	0	0	0	0.0002	0	0
Face	0	0	0	0.0016	0.0030	0	0.0002
Listen	0.0002	0	0.0002	0	0.0026	0	0
Sequence	0	0.0002	0	0.0008	0.0014	0	0.0002
Song	0	0.0001	0	0	0	0.0001	0
Song - open	0	0.0004	0	0	0	0	0
Speech	0	0	0.0006	0.0002	0.0002	0.0006	0
Sport	0	0	0	0	0	0	0

Table 3. FAR performance of each task for each participant using data from the left ear.

Task	P1	P2	P3	P4	P5	P6	P7
Breathe	0	0.0125	0	0.0125	0.0125	0.0250	0
Breathe - open	0.0500	0.0125	0.0375	0.1000	0.0375	0	0
Face	0.0125	0.0125	0	0.1125	0.4000	0	0.0375
Listen	0.0750	0.0375	0.0375	0.0500	0.3375	0.0125	0
Sequence	0.0125	0	0	0.0375	0.4000	0.0375	0
Song	0.0375	0.0125	0	0.0375	0.0500	0	0
Song - open	0.0250	0.0250	0.0500	0.0125	0	0	0
Speech	0	0.0125	0.0625	0	0.3375	0	0.0125
Sport	0.0250	0.0250	0	0.0125	0.0375	0.0125	0.0125

Table 4. FRR performance of each task for each participant using data from the left ear.

+ Examples	- Examples	FAR	FRR
P_c, T_c	P_i, T_*	0.000074	0.004424
P_c, T_c	P_i, T_c	0.000724	0.001522
P_c, T_c	P_c, T_i	0.002523	0.039702
P_c, T_c	$P_i, T_* + P_c, T_i$	0.000186	0.052565

Table 5. Four analyses in which classifiers were trained on differing negative examples paired with resulting mean FAR and FRR across all participants and tasks. P_c indicates correct participant, P_i incorrect participant, T_c correct task, T_i incorrect task, and T_* any task.

the tasks overall from most (1) to least (9) favorite. *Song - open* ranked most favorites (μ=4.25) followed by a tie between *breathe - open*, *song*, and *speech* (μ=4.75). *Sequence* (μ=7.75) and *face* (μ=6.75) were ranked least favorite.

In addition to the scales and rankings, we included a few open response questions to ascertain attitudes around use cases for in-ear EEG and passthoughts, and the comfort of wearing an in-ear EEG device in everyday life. Participants first read the prompt, "Imagine a commercially available wireless earbud product is now available based on this technology that you've just experienced. It requires minimal effort for you to put on and wear.", and were asked about use cases for in-ear EEG and passthoughts. Responses about in-ear EEG expectedly included authentication for unlocking a phone or computer and building access, but also aspects of self-improvement such as P4's response "Help people increase focus and productivity". P5 and P6 also indicated a use for measuring engagement with media like movies and music, and relatedly P4 wrote "music playback optimized for current mental state and feelings". In terms of comfort wearing such a device, participants generally responded they would be comfortable, though P5 and P6 stipulated only when they already would be wearing something in the ears like earphones. Notably, three participants also added that imagining a face was difficult

and had concerns regarding their ability to repeat tasks in the same exact way each time.

A final component of usability we acquired was the ability of the participants to recall their specific chosen passthoughts. Participants were contacted via e-mail approximately two weeks after data collection and asked to reply with the passthoughts they chose for the *song*, *sport*, *speech*, *face*, and *sequence* tasks. All participants correctly recalled all chosen passthoughts, with the exception of one participant who did not recall their chosen word component for the *sequence* task.

IMPOSTER ATTACK
While our authentication analysis establishes that passthoughts achieve low FAR and FRR when tested against other participants' passthoughts, this does not tell us how robust passthoughts are against a spoofing attack, in which both a participant's custom-fit electrode, and details of that participant's chosen passthought, are leaked and authentication is attempted by an imposter.

The first aspect of this scenario we tested was the ability of an imposter to wear an earpiece acquired from someone else and achieve viable impedance values for EEG collection based on the fit of the pieces in their ears. P1 tried on each of the other participants' customized earpieces, which were able to at least physically fit in P1's ears. The impedances of P1 wearing other participants' earpieces were then recorded and are listed in Table 6 below. While some are a better fit than others, overall they are higher (worse) than those achieved by the pieces' intended owners themselves.

To explore the scenario of an imposter attempting to gain access, we chose the case of the most vulnerable participant, P6, whose earpieces P1, P2, and P7 had the lowest impedances while wearing. We collected data using the same data collec-
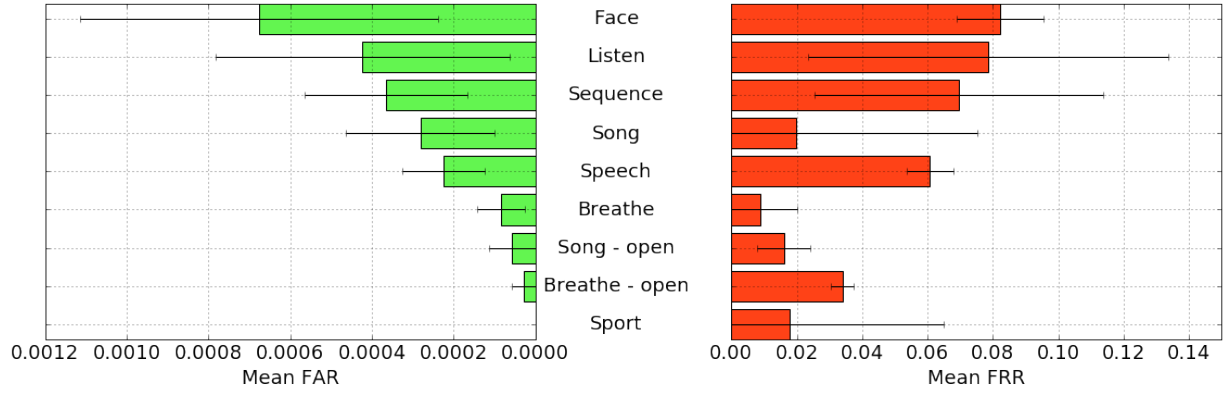
Figure 4. FAR and FRR results by task, across all subjects, using data from the left ear only.

	Impedance [kΩ]					
	Left ear			**Right ear**		
P	**C**	**F**	**B**	**C**	**F**	**B**
2	34.1	10.2	12.8	27.8	16.0	16.3
3	21.1	20.9	19.0	13.5	11.3	19.5
4	14.1	11.9	9.7	11.0	11.1	13.3
5	17.2	21.9	10.3	32.6	12.5	11.6
6	18.7	10.0	8.4	14.8	11.5	8.9
7	91.5	>1000	21.5	33.5	26.4	31.0

Table 6. Electrical impedances with P1 wearing each other participant's (P) custom-fitted earpieces, for concha (C), canal-front (F) and canal-back (B).

	Impedance [kΩ]		
P	**C**	**F**	**B**
1	18.7	10.0	8.4
2	46.7	35.7	24.8
7	44.5	20.5	26.3
X	70.0	10.5	8.9

Table 7. Left concha (C), canal-front (F) and canal-back (B) electrode impedances of "imposters" P1, P2, P7 and "PX" - a person completely outside of the system - wearing P6's left earpiece.

across all combinations, successfully rejecting the simulated imposters.

tion protocol, but had the "imposters" refer to P6's report of chosen passsthoughts. Each imposter performed each of P6's passthoughts (simulating an "inside imposter" from within the system). Following the same analysis steps, we generated 200 samples per task for our imposters, using data from all left ear electrodes.

Since every participant has one classifier per task (for which that task is the passthought), we are able to make 200 spoofed attempts with the correct passthought on each of P6's classifiers. We find zero successful spoof attempts for tasks with a chosen secret (e.g., *song* or *face*), see Table 7. In addition, we also do not find any successful spoof attacks for tasks with no chosen secret (e.g., *breathe*). In fact, in all 1,800 spoof attempts (200 attempts for each of the nine classifiers), we do not find a single successful attack on any of P6's classifiers.

However, since this participant's data appeared in the initial pool, the classifier may have been trained on his or her recordings as negative examples. To explore the efficacy of an outsider spoofing recordings, we repeated the same protocol with an individual who did not appear in our initial set of participants (an "outside imposter"). Again, we find zero successful authentications out of 1,800 attempts.

Our "leave one out" analysis can also be seen as another set of outside imposter attacks, in which each participant acts as an outside imposter for each other participant, but where the imposters have their own manufactured earpieces and passthoughts. The best task classifiers achieved FARs of 0

DISCUSSION & LIMITATIONS

Our findings demonstrate the apparent feasibility of a passthoughts system consisting of a single earpiece with three electrodes and a reference, all in or on the left ear. FARs and FRRs are very low across all participants and tasks, with FARs overall lower than FRRs, a desirable pattern in terms of authenticating access to potentially sensitive information. Participants' best-performing tasks or passthoughts typically see virtually no errors in our training. From our various training/testing schema it emerged that the inherence factor performs better on its own compared to the knowledge factor, but the combination of the two achieves the lowest FAR indicating measurable benefit of multiple factors. Passthoughts also appear to be quite memorable given our two week recall follow-up and some were rated highly repeatable and engaging. Furthermore, no spoofed attacks were successful in our cursory analyses. These results are a marked improvement upon the only other in-ear EEG passthoughts work we are aware of [8], from 80% accuracy to 99.82%. This improvement is likely due to the custom-fit quality of the earpieces, the use of three electrodes in the ear as opposed to one, and possibly the XGBoost classifier instead of a generic support vector machine.

Several tasks performed exceedingly well among participants, even tasks like *breathe* and *breathe - open* which didn't have an explicit secondary knowledge factor like in *song* or *face*. This suggests a passthoughts system could present users with an array of options for them to choose from, though it

remains to be seen how these tasks scale with larger populations. Furthermore, we were able to achieve these results by generating feature vectors based on only 500 ms (300 voltage readings across the three electrodes), which is somewhat surprising given that tasks like *song* are presumably changing over several seconds.

Our multiple classifier training strategies establish the presence of both knowledge and inherence factors as schemes isolating each still achieve low FARs. Interestingly, by training classifiers to reject both wrong participants and wrong tasks from the correct participant, thus incorporating both the knowledge and inherence factors, we see an increase in FRR suggesting multi-factor authentication's added security may come at a small cost to convenience. As preliminary evidence of the potential *secondary* inherence factor of ear shape in the current prototypes, we tested P1 using their best task (listen) with P6's earpieces but tested on P1's classifier, with a resultant FAR of 0, suggesting the protective value of the custom-fit earpieces likely due to differing electrical impedances from one user to another. To add a possession factor, these custom-fit earpieces could also easily include a hardware keypair to sign authentication attempts, functioning similarly to the way a key fob can be used to unlock a car, but compared to this single-factor system the demonstrated inherence and knowledge factors would provide important extra layers of security.

Contrary to our expectations, relatively low impedances were still achievable by imposters using another participants' custom-fit earpieces, despite the uniqueness of ear canal shapes between individuals [1]. This may have been due to our use of conductive gel and warrants further exploration. This discovery is not detrimental to the system, however, as the primary inherence factor of this system is drawn from the uniqueness of individuals' EEG patterns and not ear canal shape.

The powerful interactions between inherence and knowledge emerged in our spoofing attacks. Although our target participant documented their chosen passthought, the spoofers found ambiguity in how these passthoughts could be expressed. For the *face* task, the spoofers did not know the precise face the original participant had chosen. For the *song* tasks, though the song was known, the spoofers did not know what part of the song the original participant had imagined, or how it was imagined. This experience sheds light on passthoughts' highly individual nature and suggests there may be intrinsic difficulty in spoofing attempts. Future work should examine this effect more explicitly to elucidate the effect of knowledge task specificty on defense against imposters.

Finally, performance on Fp1 was not as high as performance in the ear, despite Fp1's popularity in past work on passthoughts [7]. This could be explained by the greater number of electrodes in the ear in this analysis (compared to one on Fp1). Additionally, Fp1 is best poised to pick up on frontal lobe activity (e.g, concentration), that our tasks did not generally involve; in fact, several involved audio (real or imagined), which we would expect to be better observed from the auditory cortex near the ears. Future work should continue to

investigate what sorts of mental tasks best lend themselves to in-ear recording.

The real-world implications of this study are limited by the controlled laboratory setting in which it was conducted as well as the small, relatively homogeneous sample of participants. In the case of this initial evaluation however, the homogeneity of our participant pool can be seen as a strength of the reported results given that system is meant to distinguish between individuals. In order to establish the validity of this system for widespread real-world use we feel it is necessary to expand the size and diversity of participants in future studies, encompassing users for whom the system would be particularly attractive such as those with extreme security needs and/or persons with disabilities which prevent them from performing other authentication methods. This sample size is comparable to that of other EEG passthoughts work [3, 17, 23, 7, 8] and other custom-fit in-ear EEG research [13, 19]. The fitting and manufacturing of custom-fit earpieces was the main limitation to increasing our sample size. This may very well pose a limitation in proliferation and adoption of such a technology as well, although recently there have been developments in at-home kits for creating one's own custom-fitted earpieces [29] that could help overcome this barrier.

FUTURE WORK

One primary question is how our passthoughts system performance will change with a greater number of users and with more diverse data. Our system specifically trains on negative examples of incorrect users; we do not yet know how this approach will scale. At the same time, we must investigate the stability of EEG readings for a passthought over time to establish long-term usability and determine when a classifier retraining period is necessary. Recent work in scalp-based BCI research has demonstrated individuals' EEG permanence over one to six months [16, 2] or even over one year [25] between sessions, though the permanence of passthougts as described here and in-ear EEG collection remains unexplored to our knowledge. Future work must also examine EEG data collected from a variety of different user states: ambulatory settings, during physical exertion or exercise, under the influence of caffeine or alcohol, etc.

Another important question surrounds how passthoughts might be cracked. Generally, we do not understand how an individual's passthought is drawn from the distribution of EEG signals an individual produces throughout the day. Given a large enough corpus of EEG data, are some passthoughts as easy to guess as *password1234* is for passwords? Future work should perform statistical analysis on passthoughts, such as clustering (perhaps with t-SNE) to better understand the space of possible passthoughts. This work will allow us simulate cracking attempts, and to develop empirically motivated strategies for prevention, e.g., locking users out after a certain number of attempts. This work could also reveal interesting tradeoffs between the usability or accuracy of passthoughts and their security.

Finally, our work leaves room for some clear user experience improvements. Future work should assess dry electrodes, commonly found in consumer EEG devices, for com-

fort and usability. Designing for a single ear from the out-set, the electrodes should be grounded to a place on that ear like the earlobe, instead of the forehead. Speakers could also be placed inside our current custom-fit earbuds to produce working "hearables" that could be used as ordinary head-phones. Future work should also attempt a closed-loop (or online) passthought system, in which users receive immedi-ate feedback on the result of their authentication attempt. A closed-loop BCI system would assist in understanding how human learning effects side might impact authentication per-formance, as the human and machine co-adapt.

CONCLUSION

As demonstrated by these preliminary results, custom-fit, in-ear EEG earpieces can provide three factors of security in one highly usable step: thinking one's passthought, using the discreet form factor of an earpiece. Among this initial sam-ple, we are able to achieve 99.82% authentication accuracy using a single sensing earpiece, showing potential for inte-gration with technology already used in everyday life (like earphones). By expanding our corpus of EEG readings (in population size, time, and diversity of settings), we hope to better understand the underlying distribution of EEG signals and security properties of passthoughts as well usability is-sues that may arise in different contexts.

REFERENCES

1. Akkermans, A. H. M., Kevenaar, T. A. M., and Schobben, D. W. E. Acoustic ear recognition for person identification. In *Proceedings - Fourth IEEE Workshop on Automatic Identification Advanced Technologies, AUTO ID 2005*, vol. 2005 (2005), 219–223.

2. Armstrong, B. C., Ruiz-Blondet, M. V., Khalifian, N., Kurtz, K. J., Jin, Z., and Laszlo, S. Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for erp biometrics. *Neurocomputing 166* (2015), 59–67.

3. Ashby, C., Bhatia, A., Tenore, F., and Vogelstein, J. Low-cost electroencephalogram (EEG) based authentication. In *2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011* (2011), 442–445.

4. Braz, C., and Robert, J.-M. Security and usability: the case of the user authentication methods. In *Proceedings of the 18th Conference on l'Interaction Homme-Machine*, ACM (2006), 199–203.

5. Chen, T., and Guestrin, C. XGBoost : Reliable Large-scale Tree Boosting System. *arXiv* (2016), 1–6.

6. Chen, Y., Sun, J., Zhang, R., and Zhang, Y. Your song your way: Rhythm-based two-factor authentication for multi-touch mobile devices. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, IEEE (2015), 2686–2694.

7. Chuang, J., Nguyen, H., Wang, C., and Johnson, B. I think, therefore I am: Usability and security of authentication using brainwaves. In *International Conference on Financial Cryptography and Data Security*. 2013, 1–16.

8. Curran, M. T., Yang, J.-k., Merrill, N., and Chuang, J. Passthoughts authentication with low cost eareeg. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, IEEE (2016), 1979–1982.

9. Frey, J. Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications. *Proceedings of the 3rd International Conference on Physiological Computing Systems*, PhyCS (2016), 105–114.

10. Garrett, D., Peterson, D. A., Anderson, C. W., and Thaut, M. H. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 11*, 2 (2003), 141–144.

11. Genaro Motti, V., and Caine, K. Understanding the wearability of head-mounted devices from a human-centered perspective. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ACM (2014), 83–86.

12. Holz, C., and Knaust, M. Biometric touch sensing: Seamlessly augmenting each touch with continuous authentication. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, ACM (2015), 303–312.

13. Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., and Mandic, D. P. A study of evoked potentials from ear-EEG. *IEEE Transactions on Biomedical Engineering 60*, 10 (2013), 2824–2830.

14. Looney, D., Park, C., Kidmose, P., Rank, M. L., Ungstrup, M., Rosenkranz, K., and Mandic, D. P. An in-the-ear platform for recording electroencephalogram. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (2011), 6882–6885.

15. Lu, H., Huang, J., Saha, T., and Nachman, L. Unobtrusive gait verification for mobile phones. In *Proceedings of the 2014 ACM international symposium on wearable computers*, ACM (2014), 91–98.

16. Maiorana, E., La Rocca, D., and Campisi, P. On the permanence of eeg signals for biometric recognition. *IEEE Transactions on Information Forensics and Security 11*, 1 (2016), 163–175.

17. Marcel, S., and Millan, J. d. R. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 4 (2007), 743–748.

18. Michalska, M. Openbci: Framework for brain-computer interfaces. *University of Warsaw Faculty of Mathematics, Informatics and Mechanics* (2009).

19. Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. EEG recorded from the ear: Characterizing the Ear-EEG Method. *Frontiers in Neuroscience 9*, NOV (2015).

20. Monrose, F., and Rubin, a. Authentication via keystroke dynamics. *Proc. of the 4th ACM Conf. on Computer and Communications Security* (1997), 48–56.

21. Nymi. Nymi Band - Always-On Authentication.

22. Peirce, J. W. Psychopy-psychophysics software in python. *Journal of neuroscience methods 162*, 1 (2007), 8–13.

23. Poulos, M., Rangoussi, M., Alexandris, N., and Evangelou, a. Person identification from the EEG using nonlinear signal classification. *Methods of information in medicine 41*, 1 (2002), 64–75.

24. Rogers, C. E., Witt, A. W., Solomon, A. D., and Venkatasubramanian, K. K. An approach for user identification for head-mounted displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ACM (2015), 143–146.

25. Ruiz-Blondet, M. V., Jin, Z., and Laszlo, S. Permanence of the cerebre brain biometric protocol. *Pattern Recognition Letters 95* (2017), 37–43.

26. Sasse, M. A., Brostoff, S., and Weirich, D. Transforming the weakest linka human/computer interaction approach to usable and effective security. *BT technology journal 19*, 3 (2001), 122–131.

27. Tartz, R., and Gooding, T. Hand biometrics using capacitive touchscreens. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, ACM (2015), 67–68.

28. Thorpe, J., Van Oorschot, P. C., and Somayaji, A. Pass-thoughts: authenticating with our minds. *Proceedings of the 2005 workshop on New security paradigms* (2005), 45–56.

29. Voix, J., Maloney, M., and Turcot, M. C. Settable compound delivery device and system for inflatable in-ear device, Aug. 18 2015. US Patent 9,107,772.