Inexpensive Brainwave Authentication: New Techniques and Insights on User Acceptance

Patricia Arias-Cabarcos
KASTEL/KIT
patricia.cabarcos@kit.edu

Thilo Habrich
University of Mannheim
t-habrich@web.de

Karen Becker
University of Mannheim
becker-karen@outlook.com

Christian Becker
University of Mannheim
christian.becker@uni-mannheim.de

Thorsten Strufe
KASTEL/KIT
strufe@kit.edu

Abstract

Brainwaves have proved to be unique enough across individuals to be useful as biometrics. They also provide promising advantages over traditional means of authentication, such as resistance to external observability, revocability, and intrinsic liveness detection. However, most of the research so far has been conducted with expensive, bulky, medical-grade helmets, which offer limited applicability for everyday usage. With the aim to bring brainwave authentication and its benefits closer to real world deployment, we investigate brain biometrics with consumer devices. We conduct a comprehensive experiment that compares five authentication tasks on a user sample up to 10 times larger than those from previous studies, introducing three novel techniques based on cognitive semantic processing. We analyze both the performance and usability of the different options and use this evidence to elicit design and research recommendations. Our results show that it is possible to achieve Equal Error Rates of 14.5% (a reduction between 37%-44% with respect to existing approaches) based on brain responses to images with current inexpensive technology. With regard to adoption, users call for simpler devices, faster authentication, and better privacy.

1 Introduction

The field of Brain Computer Interfaces (BCI) has researched and come to solutions that allow humans to communicate with machines using their brains [80]. These technologies have been especially important in the health sector, where BCIs can for example expand the interaction capabilities of people with severe paralysis [10]. But with the development of consumer-grade electroencephalogram (EEG) readers [25, 30, 33, 41, 54], new opportunities appear for using BCIs in many other realms, such as entertainment or marketing [74, 81]. Indeed, low cost headsets are already being commercialized for these purposes and we can find app stores[1] that offer brain controlled games, relaxation trainers, and several other types of applications. In this context, and further spurred by the drawbacks of using passwords for proving online identity, research on brain biometrics has recently attracted a great deal of attention.

Brainwaves – patterns of measurable electrical impulses emitted as a result of the interaction of billions of neurons inside the human brain– present particular features that make them stand out over more traditional biometrics [28, 72]. Contrary to traits like e.g., face or gait, which can be observed from the outside and potentially misused to identify users without consent [35, 78], brain activity is not observable and thus resistant to this type of surveillance. Another noteworthy aspect is that credentials based on brainwaves can be easily revoked: our brain responses vary with the stimuli, and so in the case of having brainwaves stolen, a new credential could be generated by changing its associated stimulus. Besides, given that brain activity is always present in living human beings, brainwaves can strengthen authentication with intrinsic liveness detection.

But despite the benefits of brain biometrics and the emerging democratization of EEG technology, more research is needed to make brainwave authentication applicable in real-world scenarios. Currently, the vast majority of existing work is focused on medical-grade equipment, and the scarce experiments with consumer devices involve small user samples, implement basic authentication techniques (e.g., resting), and provide limited insights on usability. Furthermore, solutions are oriented to optimize particular classification models but provide little exploration of different implementation options and their practical implications. The result is a conspicuous lack of information on how to design brainwave authentication systems for different scenarios. Motivated to fill this gap, we make two fundamental contributions to move forward:

- **(1) Design, implementation, and testing of new authentication techniques.** We focus on techniques based on the extraction of time-locked endogenous brain responses, which are known to provide higher signal-to-noise ratio than continuous EEG recordings, the common practice in related work. Apart from techniques known

[1] https://store.neurosky.com/collections/apps

in the medical-grade literature, we introduce three new tasks based on cognitive semantic processing. As a main result, we are able to achieve Equal Error Rates of 14.5%, which suppose a reduction of 37%-44% with respect to previous studies. Furthermore, we are the first to report a comprehensive comparison of brainwave authentication tasks, including testing with one-class vs two class classifiers, analyzing the relevance of features in time and frequency, considering usability, and grounded on a subject pool (N=52) that is up to 10 times larger than the sample size in previous studies.

- **(2) Usability study.** Generally, achieving high classification accuracy at the cost of low usability in authentication system design is problematic, since it can limit real-world applicability. Despite its importance, only two works so far have considered usability in the field of consumer-grade brainwave authentication. Chuang *et al.* [20] conducted an experimental user study asking participants (N=15) to rate authentication tasks according to how enjoyable, easy, or engaging they were. Besides this pioneer study, Sohankar *et al.* [66] analyzed the usability of brainwave authentication systems in the literature against an heuristic metric built on parameters such as the type of headset or the estimated time to authenticate, but without considering users' experiences and perceptions. Here, we explore the usability of the proposed authentication techniques through empirical evidence as in [20], but extending the scope of the evaluation to: 1) cover both the usability of the tasks and the brainwave device, and 2) explore attitudes towards acceptance. Our results extend and complement previous work and aid in understanding the usability-security tradeoffs to take into account when implementing an authentication system.

Apart from these two studies, we contribute to the literature by distilling lessons learned to inform future designs and research on brainwave authentication, publishing our dataset to facilitate replication and encourage further research.

The remainder of this paper is organized as follows. Section 2 introduces the status quo on brainwave authentication, defines important concepts, and sets up our application scenario. Sections 3 and 4, focus on the design of authentication tasks to collect brainwaves and detail data processing steps. We report performance and usability results in Sections 5 and 6. Finally, the paper wraps up with a discussion of lessons learned in Section 7 and conclusions in Section 8.

2 Background

To set the background knowledge for the rest of the paper, we describe here the state of the art in brainwave authentication systems, followed by a primer on their key components, and the threat model and use-case we adopt.

2.1 Related Work

Since the first human electroencephalogram was recorded in 1924 [29], many studies have shown that brain activity contains individuating patterns due to the influence of both genetic factors, e.g., given the unique folding structures of the cortex, and non-genetic factors, such as intelligence or previous experiences [11, 45, 79]. On these grounds, researchers have investigated the usage of brainwaves as biometrics for user identification and authentication. However, the vast majority of this research [28] has been conducted using medical-grade EEG equipment, which is highly precise, but at the same time expensive, bulky, and difficult to use. In this line of work, Palaniappan and Mandic [55], in 2007, recorded the EEGs of 102 subjects and applied classification algorithms demonstrating an overall authentication accuracy of 98%. This study and similar works have shown promising results and opened the door to further research with the advent of consumer-grade EEG devices in 2007. At this point, with low-cost, easy, and even aesthetic wearables, brainwave-based authentication for the masses has become a tangible possibility. And so the question arises whether it is possible to get accurate results with this type of EEG headsets.

The literature on consumer-grade EEG authentication is scarce and so far it only includes experiments with a small number of subjects[2] as opposed to the medical case. This is an important gap, since the reported accuracy may not hold when applied to larger populations where the probability of finding similar users increases [32]. Additionally, existing works mostly implement authentication based on continuous EEG recordings (e.g., while relaxing or imagining something), but few of them [2, 51, 52, 71] have looked specifically at the extraction of time-locked brain variations that appear in reaction to external stimuli. These variations, called ERPs (Event Related Potentials), have been successfully tried in research with medical EEG equipment, and they are appealing for the consumer scenario given their higher signal-to-noise ratio. Another important limitation in current research is that most publications test one authentication task but there are few comparisons between different alternatives and just Chuang *et al.* [20] have addressed the usability of EEG authentication as perceived by users, a key aspect to understand adoption.

Looking at the existing gaps, in this work we aim to move beyond the state of the art by expanding three main fronts. First, we implement new authentication tasks based on ERPs for consumer brainwave readers. Second, we thoroughly compare these tasks, evaluating not only their performance but also conducting a user study to understand usability. And third, we do our experiments on a larger set of users (N=52) and release the dataset to allow for replication and further research.

[2] Generally ≤ 10; the maximum reported number of users is 31 [2]

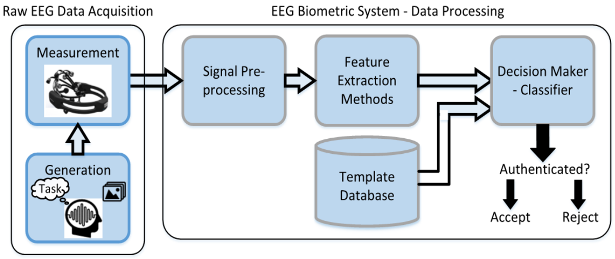


Figure 1: Structure of a brainwave authentication system

2.2 Brainwave Authentication Basics

In a biometric authentication system, users are granted access depending on their distinct physiological or behavioral traits, such as the commonly used fingerprints, voice, or face features. These traits are collected through specific sensors, processed, and compared to a previously stored sample or template from the user trying to authenticate, checking if it is a match or a mismatch. Though brainwave patterns can be used to prove a person's identity, their acquisition differs with respect to other biometrics: they need to be "generated" while performing a specific task or as a response to a stimulus, such as sounds or images. Conversely, the primary modules of a brainwave-based authentication system [28], depicted in Figure 1, are:

Generation and measurement (3). Executes the *acquisition protocol* or *task* that triggers unique brainwave activity and records the associated voltage fluctuations.

Signal Pre-processing (4.1). Treats the raw EEG signal to remove undesirable artifacts, such as interferences from nearby electronics, and increase the signal-to-noise ratio.

Feature Extraction (4.1). Isolates the signal components that are relevant for authentication, i.e., those that contain the most information about a subject.

Classification. (4.2) Implements algorithms to tell authentic and non-authentic users apart.

2.3 Use Case and Threat Model

We consider a brainwave-based authentication system that protects access to applications in a desktop or laptop computer. First, the users must complete an *enrollment phase*, where their brain signals are collected to build a classification model and stored with their identity (e.g., a username). Then, during the *authentication phase*, a user supplies her identity and receives a series of visual stimuli. The generated brain responses are compared to the stored user model for denying or granting access. Therefore, for each user with true identity ID_t and claimed identity ID_c, we test the hypotheses:

$$H_0 : ID_t = ID_c \quad vs. \quad H_1 : ID_t \neq ID_c \qquad (1)$$

to decide if the user is genuine or not (accept/reject H_0).

In this scenario, we consider a "zero effort" adversary [43]. This type of attacker tries to impersonate a valid user by claiming the target's identity (ID_u) and presenting the attacker's own biometric characteristic to the system. We assume this attacker has physical access to the device of the target victim. The resistance of a biometric system to zero-effort attacks is the system false accept rate (FAR), which we calculate, among other metrics, to discuss the performance of the proposed authentication mechanisms. We use this scenario and attacker model to guide our experiments and we further discuss the applicability to different use-cases in Section 7.

3 Brainwave Data Acquisition

In the first step of a brainwave authentication system, specific brain signals of a user need to be activated in order to generate her credential or authentication material. This process is called acquisition protocol and can be accomplished trough different types of tasks [28]. *Resting tasks*, where the user is asked to relax in a comfortable position without moving or thinking of anything in particular, are the easiest to perform. Indeed, they were among the first protocols to be investigated [60] due to their simplicity. A second category of protocols is that of *mental tasks*. In this case, users are asked to carry out imaginary actions, motor-related or not. When performing motor imaginary actions, users have to imagine kinesthetic movements of selected body parts, as opening and closing a fist or moving a finger [20]. Non-motor imaginary, on the contrary, refers to all other mental tasks that are not related to movement [83], such as mental letter composition [56], imagined speech [15], or mental calculation [48]. The last category of protocols, *stimulus-related tasks*, consists of approaches that expose subjects to stimuli of different nature (e.g., visual, auditory, emotional).

The most common approach for brainwave authentication is to use the continuous EEG signal associated to the whole duration of a task. But stimulus-based tasks offer an alternative possibility because they can also evoke specific time-locked potentials. These brain responses, called *Event-related Potentials* (ERPs) [79], appear as a temporary variation of the brainwave's voltage amplitude [36]. While more complex to implement, acquisition protocols based on ERPs provide a higher signal-to-noise ratio, being less sensitive to background perturbations [5]. This feature makes ERPs specially suitable for systems based on consumer-grade EEG devices, in which cheap sensors capture signals with lower quality compared to medical-grade electrodes [7, 24, 28]. Given the potential for ERPs to provide better accuracy, we design our tasks based on them and define the brainwave collection experiment accordingly.

3.1 Experiment Design

We focus on endogenous ERPs, a type of potentials that occur after the cognitive processing of sensory stimuli, i.e., later than 100ms after stimulus presentation[3]. While exogenous ERPs appear earlier and just depend on physical parameters of the stimulus (e.g., light intensity), endogenous ERPs are partially influenced by the subject's knowledge, motivation level, and cognitive abilities [11], and so more likely to exhibit individual characteristics useful for authentication [79]. These characteristics, together with the stable morphology of ERPs [5, 12], are the foundations for the uniqueness of this type of brainwaves. The most relevant **endogenous ERPs** are the P300 and the N400:

P300. It is a positive wave that peaks around 300ms after exposure to a certain stimulus [36]. This wave is triggered if a subject decides consciously or unconsciously that a presented stimulus or event is rare. In experimental setups, a P300 response can be elicited using the *Oddball Paradigm* [67], in which low-probability target items (e.g., pictures) are mixed with high-probability non-target or "standard" items.

N400. It is a negative wave that peaks at 400ms after a stimulus [38]. While the P300 is related to the attention of a subject, the N400 appears related to tasks that require semantic processing [36], such as language processing.

We devised five acquisition protocols to elicit the described potentials for authentication. The first two protocols focus on the P300 ERP, and were selected based on their successful application with medical-grade equipment. Besides, to further explore the space of possibilities, we introduce three new tasks built on the N400 potential that have never been used for authentication. The following list describes how we implemented the **acquisition protocols** grounded on neuroscience research techniques to trigger ERP potentials [23, 36–38, 67] :

P300:Selected. This task elicits the P300 potential based on the *oddball paradigm*. We first let the user pick a picture of her choice, which will be the target stimulus. The authentication task consists of looking at a sequence of images where the target image appears infrequently. Upon appearance, because it is a rare occurrence, a P300 is evoked that differs across subjects. To increase the attention and therefore the wave amplitude, we instructed the users to count the occurrences of the target stimuli.

P300:Assigned. Same as P300:Selected, but the is assigned the rare image.

N400:Words. This task is based on a *semantic priming paradigm*. Priming is defined as *"an improvement in performance in a perceptual or cognitive task, relative to an appropriate baseline, which is caused by previous, related experience"* [73]. Simply put, a subject is primed on an object if it has previous experience with this object. After priming, if the subject is presented with a semantically related stimulus, the brain finds it more meaningful and so the N400 potential appears. In our experiment, subjects watch a 'priming video' that displays cars driving on a highway. Afterward, several words are shown on the screen. A minority of these words is strongly related to the priming objects and aims at triggering N400 responses, and the rest are randomly generated.

N400:Sentences. This task is based on the concept of *incongruent sentences*. The N400 has been proved to appear when subjects read sentences word by word that end in a semantically incongruent manner [37]. An example for such a sentence is: *"Steve sat down to eat his* ***car****"*. Furthermore, the amplitude of the N400 wave depends on the subject's expectancy for the final word. This means that if subjects are primed on certain congruent endings, the N400 response is stronger when the incongruent word appears [38]. We therefore base on this observation to build our experiment. The task consists of showing users a sequence of sentences with slight variations. First, the sentences have semantically congruent endings, but the last variation finishes with an incongruent word to elicit a strong N400.

N400:Faces. This task is based on the concept of *inhibition of knowledge* associated to N400 potentials evoked during face identification, which is another type of cognitive semantic processing, different from words. Previous work has determined that the amplitude of this wave is stronger when looking at an unfamiliar face after being presented (and therefore primed) with a sequence of familiar faces [23]. The reason is that when seeing familiar faces, the brain activates semantic representations useful to cognitively process and identify them, but these representations need to be removed and new ones activated when we start to process a new and unfamiliar individual. This inhibition of knowledge intensifies the N400. On these grounds, our protocol shows unfamiliar faces within sequences of likely familiar faces (celebrities).

3.2 Experiment Execution

Goal and Structure. The experiment at the core of this research has two goals: 1) eliciting and recording ERPs with individuating features to be used for authentication; and 2) collect information on the perceived usability of brainwave authentication. Figure 2 illustrates the brainwave collection part of the experiment, based on the five acquisition tasks described in Section 3.1. After providing consent to take part in the study, participants were told to sit comfortably and move as little as possible during the experiment. Every room was kept rather dark and quiet, in order not to disturb the subjects. Next, their brainwave activity was recorded while performing the authentication tasks. As shown in Figure 2, the recording starts with baseline measurements of brain activity while rest-

[3] A comprehensive overview of currently known ERPs identified in neurological research can be found in [69].

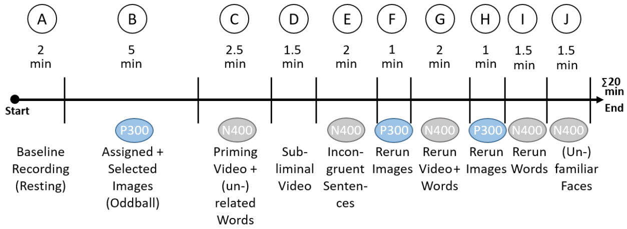


Figure 2: Graphical flow of the experiment tasks to record users' brainwave activity for authentication. Each task is briefly described, labeled with the potential meant to be evoked (P300 or N400), and tagged with its duration.

ing. Then, it follows with several sequences and repetitions of the authentication tasks[4], to acquire multiple samples for training and testing the classification algorithms. After the recording, participants filled out a paper questionnaire to assess the usability of a brainwave authentication system based on the performed tasks and headset (details in Section 6). All experiment materials are linked in Appendix 8.

Apparatus. We use the Emotiv EPOC+ headset [25] to record brainwave activity. We chose this device because it is the prevalent choice in scientific studies and it offers a higher number of recording channels (14) than other consumer grade products, which leads to more accurate measurements[5]. The experiment flow was programmed with PsychoPy [58], an open source tool for conducting experiments in behavioral sciences, and connected to the EPOC's reading software to synchronize stimuli presentation with brainwave recording.

Recruitment and Ethical Aspects. We recruited participants following a self-selection sampling approach [39]. The study was advertised through different channels asking for volunteers, including online posts, flyers spread at different university locations and brief announcements during lectures. Each participant received information about the experiment and about how we would treat their personal data fulfilling the EU General Data Protection Regulation (GDPR) [26], in order to get informed consent. To avoid biasing the subjects, we disclosed the actual purpose of the experiment, i.e., building an authentication system, at the end of the recording session and before the usability questionnaire. The approximate average duration of the whole study was 45 minutes and we compensated participants with 5€ and a report on their brainwaves containing information about interest, stress, and focus level during the study. Subjects were also told that participation was voluntary and the experiment could be abandoned at any time. The whole procedure is IRB-approved.

Participant Demographics. In total, 56 subjects took part in the experiment, conducted between May 8 and July 2, 2019. We recorded ERPs from 23 females (41.1%) and 33 males (58.9%), leading to an slightly imbalanced gender distribution. With regard to age, our population is skewed towards young adults because most of the experiments were conducted with university students. The majority, 28 subjects (50%), fall in the age range 18-24, followed by 16 (28.6%) participants aged between 25 and 31, and 8 (14.6%) in the range 32-38. The remaining 4 persons (7.2%) were over 39 years old.

4 Brainwave Data Processing

Before we can get useful brainwave data for the classification algorithms that implement authentication, raw EEG signals must undergo a two-step preparation process to: 1) remove undesirable artifacts, and 2) extract relevant features for authentication. This section summarizes the data preparation steps, following common practices in the literature [28], and the classification models we apply to these data.

4.1 Pre-processing and Feature Extraction

The data recorded during the experiment contains continuous EEG measurements of about 20 minutes length, captured at a sampling rate of 256 Hz. However, only specific relevant sections around the presentation of stimuli, i.e., the ERP waves, are required for authentication. These sections are also called *epochs* and constitute a user sample. To extract the ERPs, we cut 1-second length epochs from 100ms before stimulus presentation until 900ms thereafter to guarantee that we get the potential's information, considering variances in the peak latency [69]. After epoch extraction, we filtered electrical noise and removed samples with bad quality measures or containing large artifacts that contaminate the EEG signal (e.g., eye or muscle movements). With the clean EEG signal, the next step is to obtain discriminant features that represent and encode the mental activity of a user [28]. We chose the most common features in the *time* and *frequency* domains applied in previous works [1, 3, 6, 28, 82], and used them as a basis to further identify which features work best for our proposed tasks (see Section 5.2.3). First, considering the ERP epoch a 1-second time series, we fit it to an *Autoregressive* (AR) model with 10 coefficients and take them as features. Second, we split each 1-second epoch into five segments of 200ms and calculate their Power Spectrum (PS) in different frequency bands (α [10-13Hz], β [13-30Hz], and γ [30-50Hz]). Moreover, we generated 15 cumulative features by aggregating the PS of all 14 channels per segment, and 3 highly aggregated features, by grouping the PS of all segments into one feature per frequency band. Table 1 shows the final datasets after pre-processing, linked in Appendix 8.

[4] Element D in the study flow depicted in Figure 2 was included to test subliminal manipulations Since we did not obtain conclusive results in this regard, we just report it as a study item without giving further details

[5] The reader is referred to [65] for a comprehensive review and comparison of consumer grade EEG readers, including research applications

Dataset	#users	#samples
P300:Selected	52	911
P300:Assigned	52	910
N400:Words	52	1733
N400:Sentences	50	276
N400:Faces	50	424

Table 1: Brainwave datasets for five authentication tasks.

4.2 Classification

For the purpose of authentication, the recorded data samples of each user need to be compared to stored samples of the same subject and classified as matching or not. We compare and discuss the applicability of two authentication model approaches: 1) *one-class classifiers* (aka anomaly detectors), which only require training data from the genuine user; and 2) *two-class classifiers*, which are trained on data from both authentic and impostor users. For each category, we chose a small set of representative approaches suited for our dataset dimension, namely:

One-class classification. We implement a *k-Nearest-Neighbour* (*kNN*) method to classify users based on distance to training instances, and a one-class *Support Vector Machine (SVM)*.

Two-class classification. We chose a probabilistic *Gaussian Naïve Bayes (GNB)* classifier, and the two most common linear algorithms, Logistic Regression, and linear *Support Vector Machines (SVM)*.

We refer the interested reader to related work for more details on these models and their applications [28, 40].

5 Authentication

This section evaluates the performance obtained for the proposed authentication tasks, comparing one-class vs two-class classification algorithms, analysing feature relevance, and contextualizing the results with regard to related work.

5.1 Evaluation Metrics

Several methods can be applied to evaluate classification systems. In the case of a binary problem, there are four possible classification results: 1) authenticate a legitimate user (True Positive or TP), 2) authenticate an illegitimate user (False Positive or FP), 3) deny an illegitimate user (True Negative or TN), and 4) deny a legitimate user (False Negative or FN). Based on the frequency counts of these results, the performance of the system is typically assessed by its False Acceptance Rate (FAR), False Rejection Rate (FRR), and Accuracy (ACC). The FAR compares the number of false positives to the sum of false positives and true negatives, i.e., how often an impostor is authenticated as legitimate. In turn, the FRR compares the number of false negatives to the sum of true positives and false negatives, giving an idea of the frequency at which the system rejects legitimate users. Finally, the ACC represents the number of correct predictions over the total number of predictions made by the classifier. These metrics, however, are tied to a specific configuration of the classification threshold. Instead, we visualise results with Receiver-Operating-Characteristic (ROC) curves, which plot the FAR and True Positive Rate (=1-FRR) as a parametric function of the threshold. We also report Equal Error Rates (EER), as a summary metric that represents the point where FAR and FRR are equal. This reporting scheme, as suggested by *Sugrim et al.* [68], allows for a better understanding of the operation capabilities of authentication methods, and how they can be configured for different use-cases.

5.2 Results

We evaluated user authentication for the five defined tasks using one-class and two-class classifiers. We remove users with less than 5 samples from the datasets to have enough data for training. The one-class SVM (with Radial Basis Function kernel) and kNN (k=2) classifiers were trained on the samples of one single user[6], considered the legitimate user, and then tested with samples from both the legitimate user, which should be recognised based on the learned model, and all the other illegitimate users, which should be rejected as outliers. For two-class classification, we followed a *one-vs.-all* approach [61]. According to this scheme, we built specialized classifiers per user by assigning all the samples from this user with the “authenticated” class label, and all the others with the “rejected” label. We applied grid search to select the best features (based on their statistical significance to classify the authentic user) within a nested stratified 5-fold cross validation loop[7]. For every classification algorithm, we run the evaluation process for all the users in each dataset and we report the average EERs.

5.2.1 One-class vs Two-class Classifiers

The overall results are summarized in Table 2. As expected, the performance with two-class learning is better than that of one-class classifiers. Binary classifiers are usually more powerful, since they characterise the legitimate user in contrast to others, whereas anomaly detectors can only check for deviation from the legitimate user’s behaviour. In practice, this means that a set of anonymous user’s data needs to be pre-loaded in the application or device that offers brainwave authentication. Then the classification model can be realized

[6] We used a split ratio of 0.6 to 0.4 for training and testing sets

[7] 5-folds in inner and outer loops

	Equal Error Rate (%)				
	One-Class		Two-Class		
Task	kNN	SVM	GNB	LG	SVM
P300:Selected	49	44	**24.89**	30.85	33.5
P300:Assigned	49	42	**23.45**	30.53	34.14
N400:Words	49	40	**21.21**	30.21	31.22
N400:Sentences	48	43	**20.34**	26.14	29.31
N400:Faces	47	40	**14.5**	30.21	32.76

Table 2: Average Equal Error Rate (EER) for five authentication tasks comparing one-class vs two-class classifiers.

by combining the data of genuine users. While this type of implementation is feasible and has been proposed for other behavioral biometrics [17, 70], further research is needed on how to anonymize brainwave data.

5.2.2 ROC-based Performance of Authentication Tasks

Overall Performance. With regard to authentication tasks, our results establish the N400 protocols as better authentication options than the P300 protocols, and the best performing task is the N400:Faces, with an average EER of 14.5%. Fig. 3 shows the ROC curves for the best classifiers, illustrating the operational range of the five authentication models. The area under the curve (AUC) represents the probability that a random illegitimate user is scored lower than a random genuine user, i.e., how well the classifier can separate users. Looking at these metrics, while the N400:Faces outperforms the rest of the tasks in the tested conditions, all schemes show potential for discerning users and could therefore be feasible for brainwave-based authentication. However, there is a high variability from the average ROC curves. In this regard, an important factor to consider in the comparison is the different number of samples and users per task. As it can be observed in Fig. 3a, the N400:Words task has the highest number of samples (1730 for 51 users), which almost doubles those available for the P300 tasks (911 and 910 for 52 users). In the case of the remaining N400 protocols, the datasets are reduced to 33 users and 198 samples for the N400:Sentences and 44 users and 406 samples for the N400:Faces. Accordingly, it can be observed that protocols with less users perform better, which can be related to a higher probability of having similar users in the datasets or having more users for whom the acquisition process failed to achieve brainwave data with good enough quality. In the case of N400:Sentences, the performance can be negatively impacted by the low number of samples per user (6 on average), which leads to very few data for training, testing, and validation.

Applicability. In real world authentication scenarios, systems do not operate at the EER, but at configuration points were the FAR is lower than the FRR, to minimize the probability of impostors accessing the system. In general, most biometric systems have a FRR ranging from being falsely rejected one out of five times up to one for every thousand times (i.e., 20% to 0.1%) [34]. The FAR is more critical for security and usually ranges from 1%, for low security applications, to 0.00001% for very high security applications [18]. In this sense, our ROCs show that authentication based on the N400:Face task can be configured for best accuracy at a FAR of 1.8% and associated FRR equal to 46%. While the FAR value is close to the needs of low-security application scenarios, the FRR is unacceptably high. This same trend is observed in the ROCs for the rest of the tasks. However, we expect lower error rates in real implementations with personalized stimuli. We measure and report the FAR calculated by directly comparing impostors' ERP samples to the legitimate user model. But if we consider the dynamics of the authentication protocols, those ERPs should appear in response to the target stimuli (e.g., unfamiliar faces within a series of familiar ones). Checking this condition before accepting an ERP will yield lower FARs, as it is highly unlikely that an impostor reacts to the stimuli designed for the legitimate user. Therefore, the obtained FAR is to be understood as a rough upper bound.

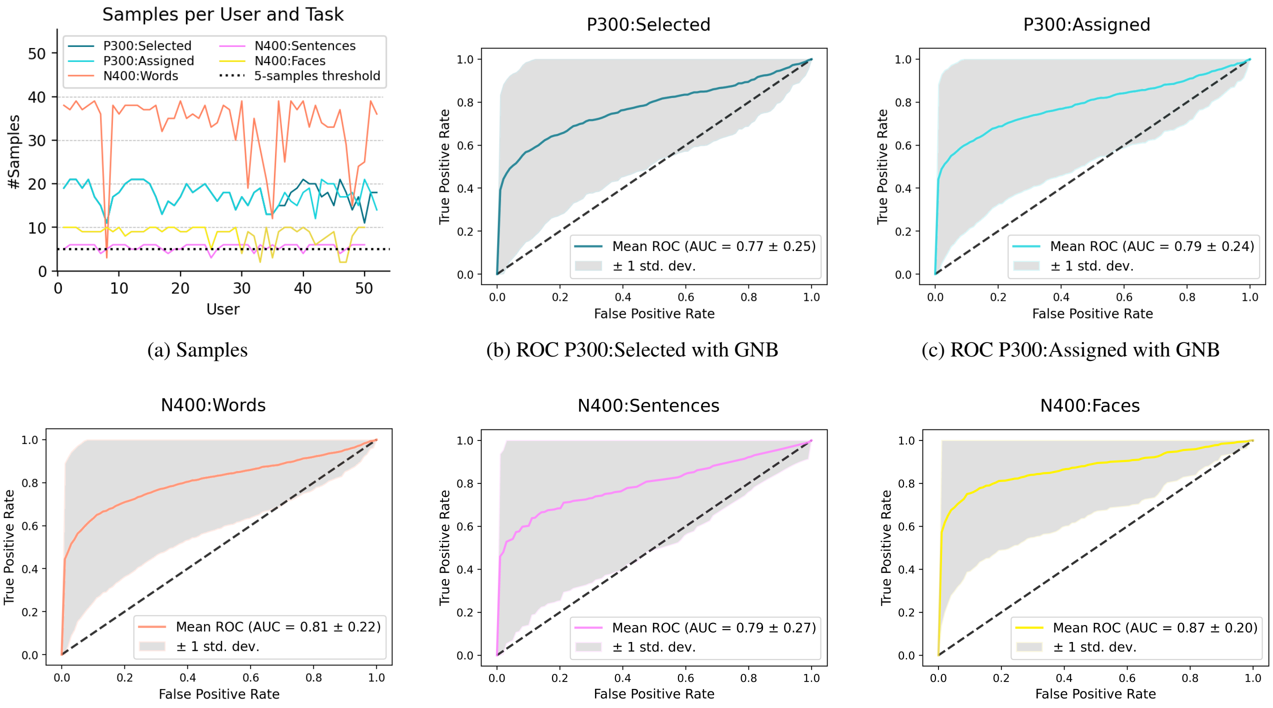
5.2.3 Feature Relevance

In addition to classification performance, we analyzed the importance of the features for classification to inform future designs of brainwave authentication prototypes. Fig. 4 shows a heatmap of selected features across the different user classifiers for the five authentication tasks. The most commonly removed features are located in the α frequency band. This is reasonable, since brainwaves in this band are the most dominant rhythm and correlate with mental states of no attention, being stronger when the eyes are closed [4]. They have been proved useful in brainwave authentication based on relaxation tasks [50], but are not applicable for the tasks proposed in this paper. Instead, the β and γ waves are usually exhibited in states of focused attention and active information processing [1], which can be the reason why they are more relevant for classification in our visual and semantic processing tasks.

5.2.4 Comparison with Related Work

Performance. Comparison with existing works on brainwave authentication is challenging due to the frequent underreporting of metrics (usually presented for an optimized configuration without providing ROCs) and the differences in the number and diversity of samples, algorithms, experimental conditions, and other aspects that influence performance. Acknowledging these difficulties, we first compare against systems[8] using consumer-grade EEG readers that report EERs, and then, to broaden the comparison, we contextualize our results with regard to other relevant works in the literature.

[8] Excluding multi-modal and multi-factor authentication approaches



(a) Samples

(b) ROC P300:Selected with GNB

(c) ROC P300:Assigned with GNB

(d) ROC N400:Words with GNB

(e) ROC N400:Sentences with GNB

(f) ROC N400:Faces with GNB

Figure 3: Performance comparison of five authentication tasks using Gaussian Naïve Bayes (GNB). Fig. (a) shows the number of samples per subject and task available for classification, using a minimum threshold of five samples per user. Figs. (b), (c), (d), (e) and (f) depict the ROC curves for each authentication task.

Nakanishi *et al.* investigated various authentication tasks [47, 49, 51–53], including resting (EER=11%, n=23), driving (EER=22-24%, n=10-30), low intensity visual stimuli (EER= 23%, n=20), and ultrasound stimulation (EER=26.2%, n=10). In all cases, our N400:Faces protocol has better or similar performance[9]. Furthermore, when compared to the ultrasound and visual tasks, which are based on ERPs and therefore closer to our proposal, we decrease the EER from 23%-26.2% to 14.5% for the N400:Faces task, which means a relative error reduction of 37-44%. These results indicate that visual tasks based on cognitive semantic processing are more suitable for brainwave authentication than current ERP-based proposals in the literature. The only other works reporting lower EERs use multi-modal fusion [6, 52] (EER=4.4% and EER=0%) or a second factor [2] (EER=0.89%) to complement brainwaves, which suggests these are viable paths to further improve the applicability of our tasks.

Though not reporting EER, the study by Chuang *et al.* [20] is specially relevant because they get high authentication accuracy using a 1-electrode EEG reader. Their best performing task is moving a finger, with FAR=4% and FRR=76% (n=15). But applying customized thresholds per user, they move up to a 0% FAR and FRR=9% using a mental singing task for authentication. If we apply a simple threshold selection (maximizing TPR-FAR) to the N400:Faces protocol, our performance also improves, achieving a point where FAR=8.5% and FRR=10.4%. This is a good result for practical applicability, considering that the FAR is already an upper bound (see Section 5.2.2), and we expect even better performance with more personalized thresholds and additional optimizations.

Looking at the literature using medical-grade EEG readers, the work by Das *et al.* [21] is the closest to ours. They use P300 ERPs for authentication, achieving EERs around 13% (n=50) with 17 sensors. We show that it is possible to achieve comparable results with N400 potentials and a simpler headset. There are also relevant studies demonstrating the value of ERPs for biometric identification, such as CEREBRE [62], which provides 100% accuracy in identifying 52 users. Though not directly comparable, it provides interesting insights on how to optimize classification through voting schemes, which could be also applicable to improve performance on the authentication case.

Participant Pool Size Considerations. The ISO-19795 [31] for biometric testing recommends 300 samples (as a minimum lower bound) for 95% confidence on a FAR <=1%. We targeted approximately this minimum size in our datasets,

[9] We computed the variation of EER with the number of subjects for the implemented tasks at points n={5,10,15,20,25,30,35} and use the closest EER value when comparing with related works tested on a smaller sample. Subjects were randomly selected and the EER averaged across 5 repetitions.

following also the recommendation that the participant pool should be as large as practicable. Our final pool size, 52 users, is bigger than that used in previous works with consumer-grade EEG readers, which implies more reliable results. Results on small datasets can be over-optimistic due to chance in the selected participants, but statistical confidence increases with more users and samples. We experimentally observed[10] that as the participant pool size increases, the variance of error estimates decreases. For example, when testing the N400:Faces for 5 users, we got an average EER=9.23% and standard deviation σ=7.7%, observing EERs as low as 2%. But the error stabilizes as the number of participants grow, getting to an average EER=14.38% and σ=0.72 at 40 subjects. We therefore contribute to understanding the uniqueness of brainwaves at a larger scale, with higher confidence. One of the main open challenges that follows from here is scaling up to bigger populations, given that the minimum sample size recommended to test for a FAR of 1:100000 is 300000 samples. As a first step towards real prototypes, our results and discussion show practicality and can help inform the design of future authentication systems.

6 Usability

This section describes the user study conducted to evaluate usability aspects, reporting quantitative and qualitative results.

6.1 User Study Design and Methods

Design. Each person taking part in the overall authentication experiment was asked to fill out a usability questionnaire that includes three categories of questions. First, we explore the perceived *usability of the five authentication tasks* asking if they are boring, require attention, and are appealing to repeatability on a daily basis. These questions are taken from Chuang's *et al.* work [20], though we ask for ratings on a 5-point Likert scale to allow for more granularity in the responses. Second, also on a 5-point Likert scale, we question about *device usability*, considering two dimensions: ability to set up the device and overall usage experience. Third, we target *acceptance*. Inspired by the work of Payne *et al.* [57] on the acceptance of tokens as authenticators, we include two open-ended questions about potential problems (Q1) and suggestions for improvement (Q2) of the brainwave authentication concept. Note that users do not evaluate a prototype but the proposed authentication tasks and the perception of how an hypothetical brainwave-based system built on these tasks would work for them in daily life. The nature of the study is therefore exploratory and oriented to inform prototype design, whose evaluation would require further testing.Thus, we cannot use the Standard Usability Scale (SUS) [16] and other well-established usability metrics (speed , error rate) applied in authentication research [22, 64], as they are only appropriate for testing prototypes with (at least) moderate functionality.

Analysis. Usability questions elicited responses on Likert scales that we analyzed with the Friedman test for omnibus comparisons. Post hoc analysis with Wilcoxon signed-rank tests were conducted with a Bonferroni correction applied, to determine which authentication tasks differed significantly. As for the open-ended questions on user acceptance, we analyzed the responses following an iterative, inductive coding approach [46]. One member of the research team read responses and created the codebook with thematic codes (see Appendix 8), and a second researcher independently coded the full set of data. The inter-coder reliability for the final codes was satisfactory for both questions: excellent agreement for Q1 (Cohen's kappa=0.91) and substantial for Q2 (Cohen's kappa=0.76). The cases where the coders differed in their final codes were discussed and reconciled.

6.2 Results

All 56 subjects replied to the Likert-ranked questions about the usability of authentication tasks and device. With regard to the open-ended questions, 28 subjects named potential problems, and 45 reported improvement suggestions for a brainwave authentication system. Here we analyze these data, providing representative user quotes when meaningful.

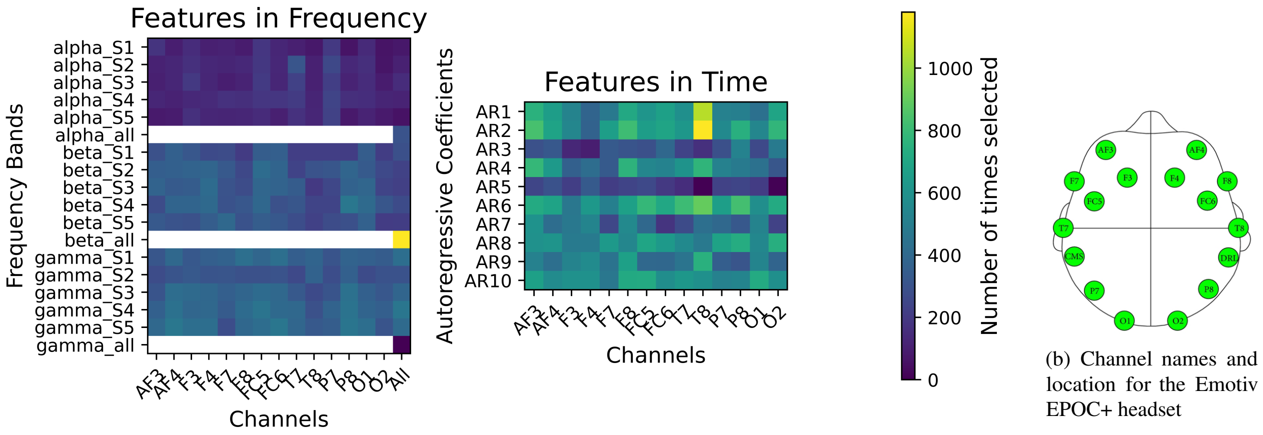
6.2.1 Perceived Usability

Usability of the Authentication Tasks. The graphs in Fig. 5 show participants' answers about tasks' usability. Answers to "boring" and "required attention" were coded from Strongly Agree (SA)=1 to Strongly Disagree (SD)=5, and answers to "Repeatability", from SD=1 to SA=5. Therefore, higher values always indicate more positive evaluations.

Analyzing the responses regarding *boredom*, protocols were rated differently ($\chi^2(4)$=108.864, p<.05). More specifically, there were statistically significant differences (p<.01) in all cases except between the P300:Assigned and P300:Selected, and the N400:Sentences and N400:Faces. The N400:Words protocol received the lowest grades with a median rating of 3 (μ=2.95, σ=1.21). With slightly better grades, the P300:Selected (μ=3.46, σ=1) and P300:Assigned (μ=3.39, σ=0.93), received a median of 3 and present no statistically significant differences. At the other extreme, the N400:Faces protocol (μ=3.78 , σ=0.99), and the N400:Sentences (μ=3.71 , σ=0.97), with the same median rating of 4 and no statistically significant difference, got the best evaluations. About the latter, one of its positive aspects is that the sentences were unexpected and sometimes funny, which makes the task more engaging, as this participant put it in the open-ended answers:

> *"I like the idea with incongruent sentences. Generally, I think that it is important to include something*

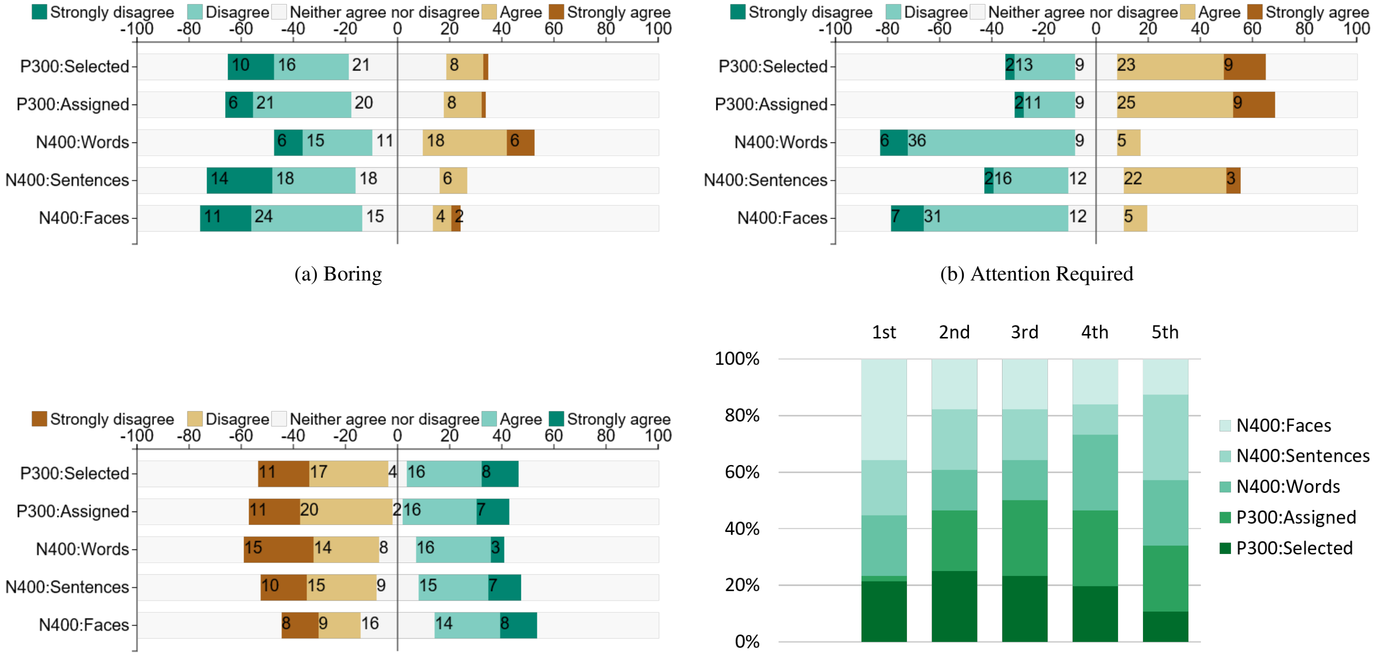
[10] 5 random selections of subjects for each participant pool size tested



(b) Channel names and location for the Emotiv EPOC+ headset

(a) Heatmap

Figure 4: Fig. (a) shows the heatmap of selected features for the GNB classification algorithm across five brainwave authentication tasks. Frequency features are calculated as the Power Spectrum of the user ERP signal in segments (S1-S5) of 200ms for the α, β, and γ bands. The time features are 10 Autoregressive Coefficients of the ERP. Features are obtained at 14 measurement channels, whose corresponding electrode positions in the scalp are depicted in Fig. (b). CMS/DRL are reference electrodes.



(a) Boring

(b) Attention Required

(c) Repeatable on a daily basis

(d) Enjoyability Ranking

Figure 5: Participant answers to the statements: (a) *"The task was boring"*, (b) *"The task required a lot of attention"*; and (c) *"I could imagine to perform this task on a daily basis at a PC for authenticating"*, for the five implemented authentication tasks. Sub-figure (d) shows how respondents ranked the tasks depending on enjoyability.

funny or encouraging to avoid boredom". (P28)

When it comes to *required attention*, tasks were also rated differently ($\chi^2(4)$=158.501, p<.05). Statistically significant differences (p<.01) appear in all cases except between the P300 protocols and the pair N400:Faces-N400:Words. The protocols with lower grades are the P300:Assigned (μ=2.5, σ=1.09) and the P300:Selected (μ=2.57, σ=1.13), both with a median of 2 and no statistically significant differences. Participants rated the attention demand of the N400:Sentences task (μ=2.85, σ=1.03) slightly better, with a median of 3. But the highest rates were assigned to N400:Faces (μ=3.73, σ=0.8) and N400:Words (μ=3.77, σ=0.76), both with a median of 4 and no statistically significant differences.

The responses regarding *envisioned daily usage* show differences too ($\chi^2(4)$=62.254, p<.05), but they exhibit a smaller variance compared to the prior questions. In this case, N400:Faces (μ=3.09, σ=1.27), with a median of 3, is the task for which most subjects reported to "strongly agree" that they would like to perform it on a day-to-day basis. In turn, the N400:Words (μ=2.61, σ=1.3) got the worst evaluation, with a median of 2. The rest of the authentication tasks fall in the middle. Statistically significant differences (p<.01) appear in all cases except between the P300 protocols, and between P300 and N400:Sentences.

Finally, when we asked participants to rank the authentication tasks, the most enjoyable protocol was the N400 Faces, chosen by 36% (20) of the respondents. At the other end of the rank, the N400:Sentences task was selected as the least enjoyable by 30% (17) of the participants. Overall, image-based tasks are preferred over text-based ones, as it was also recalled by several participants in the open-ended questions:

> *"Picture recognition is better than text recognition".* (P22)

Usability of the EEG Device. Most of the participants (62.5%) think they will be able to put on the headset by themselves, while only a 21.5% (12) reported that they do not imagine themselves completing the device setup. A plausible reason for this 21.5% could be that the headset setup required several minutes in some cases, where the hair density between the electrodes and the skin was thick. Nevertheless, the experience using the headset was mostly rated positive, with a 59% (33) of participants agreeing or strongly agreeing to this perception and no reported strong disagreements. These results indicate that authentication using the EPOC+ headset could be accepted (positive experience) but the usability of the device can still improve.

6.2.2 Attitudes towards Acceptance

Problems. Participants identified issues related to the *brainwaves* (28%), the *device* (22%), and the overall authentication *system* (50%). First, users reported concerns about the uniqueness of brainwaves and their stability against e.g., emotional influences due to stress or illness. They were also worried that familiarization with the stimuli would result in weaker brainwave responses and lead to authentication errors. Besides, one subject wondered if not being fully attentive, or as he/she put it *"having meandering thoughts"*, would affect authentication. Second, the negative points about the device were the cost, its design, and the complex setup process. Similarly, users highlighted the technical problems, such as the imprecision of the sensors. Third, participants criticized aspects of the system as a whole, specially its performance (authentication speed), usability, and the level of security and privacy provided. As illustrated by the following sample answers, users are worried about the strength of this type of authentication against attacks (even mind manipulation) and about the usage of brainwaves to infer sensitive personal information.

> *"Skepticism of the user regarding data security and other aspects which could be figured out about the users, which the user does not want.".* (P9)
>
> *"Changing of individual opinion due to presented stimuli, e.g., in particular politicians".* (P41)

In the usability category, the inclusiveness of the brainwave authentication system was the most frequent topic. Participants remarked that using sentences as stimuli would not work to authenticate children and that the system might not be usable for people with different cognitive abilities.

Suggestions for improving. Participants reported ideas that fall in three categories: *device* improvements (18%), *protocol* improvements (39%), and *system* improvements (42%). Regarding the device, users pointed to different designs that blend more naturally with everyday life, such as integrating EEG readers within headphones or hats. Another frequent comment was the need to reduce the number of electrodes and make the device simpler and easy to handle. Regarding the improvement of protocols, subjects expressed a preference for visual stimuli vs textual stimuli and call for authentication tasks that are enjoyable or "cool". As alternative tasks, for example, two participants mentioned that they *"would be interested in authentication using music or tones".* In the last category of suggestions, targeting the overall system, performance was the most frequent concern. Users suggest to *"Keep the authentication process as short as possible"*, because otherwise *"one sees the repeated, three second long typing of a password as less annoying than performing one of these [brainwave authentication] tasks as a whole"*. The effort, as stated by one of the respondents *"needs to be adapted to the required security level"*.

7 Discussion

Here we report lessons learned when designing protocols for brainwave authentication, report security considerations, and discuss practical implementation aspects and limitations.

7.1 Protocol Design

Design Effort. We argued in Section 5 that one potential reason influencing the performance and comparability of the authentication protocols was the different available number of samples for training the models, which, in our study, was affected by the protocol design effort. The number of epochs usable for classification is limited by the total number of target stimuli, i.e., those that generate an ERP, presented during the experiment. As summarized in Table 3,

Design Aspects	P300:Selected	P300:Assigned	N400:Words	N400:Sentences	N400:Faces
Avg. time[a] between target stimuli (s)	6	6	4.15[b]	14	6
# Target stimuli per round	6	6	13	6	10
# Protocol rounds	3	3	3	1	1

Table 3: Design aspects of brainwave acquisition protocols

[a] Rounded

[b] Plus the duration of the preceding priming video (24s in our experiment)

Criteria	P300:Selected	P300:Assigned	N400:Words	N400:Sentences	N400:Faces
Accuracy	- -	- -	-	-	+ +
Boredom	+	+	-	+ +	+ +
Required level of attention	- -	- -	+ +	-	+ +
Daily Usage	-	-	- -	-	+
Enjoyability	+	-	+	- -	+ +
Elicitation effort	+ +	+ +	+	- -	- -
Stimulus reusability	+ +	+ +	+	- -	- -

Table 4: Overall comparison of authentication protocols

both the N400:Sentences and N400:Faces have less total stimuli in comparison to their counterparts. There are two reasons for this: highest elicitation effort (more time required for stimuli presentation) and low stimuli reusability. While it is rather quick to present new stimuli in the N400:Words, N400:Faces, and P300 protocols, that was not possible in the N400:Sentences. In this case, the subjects first had to be primed on the congruent form of a sentence and then later on shown the incongruent version to obtain the desired ERP in response. This process takes about 14 seconds per sentence in total, which results in a smaller number of stimuli per minute. Furthermore, the incongruent sentences need to be altered each time, otherwise they would not appear incongruent to the users anymore after a small number of iterations. Similarly, the N400:Faces also suffers from this effect, i.e., an unknown face would not lead to the same reaction if it was shown repeatedly. Because of the lack of stimuli reusability, we limited the execution of these protocols to just one round in our experimental setting, with the consequential decrease in the number of samples. In the N400:Words protocol, a video and the associated words can be used several times, since only the interaction between the words and the video are important. But the best design case is that of the P300 protocols. Here, the stimuli can be endlessly reused because the brain reaction responds to an infrequent event, the oddball, but it is not related to the semantic processing and so unaltered by stimulus familiarity.

Overall Protocol Comparison. We provide a comparative summary of the analyzed protocols to inform the design of future brainwave authentication systems (see Table 4).

Considering classification performance, the N400:Faces task is the best option. This performance, combined with the highest usability scores of all tested tasks, makes it a suitable candidate for real-world implementations. The main negative aspect is the complexity of the protocol design. Thus, research towards facilitating this design process is desirable. The second best option in terms of accuracy are the remaining N400 protocols. In this group, the N400:Words shows better potential for applicability, given its higher usability results on enjoyability and required attention, as well as the lower design effort with respect to the N400:Sentences. The P300 category of protocols showed the worst accuracy. In this case, usability improves when users select their own secret image. This preference on active selection was also observed by Chuang *et al.* [20] in protocols where users either had to chose or were imposed a mental task for authentication. The most positive of P300 protocols is that they are the easiest to implement.

In summary, N400:Faces was the most accurate task and the best ranked by users, outperforming the rest of the protocols in all dimensions. Nevertheless, performance needs to be further improved for its application in real scenarios.

7.2 Security

In this paper we covered a zero effort attacker model, but, like in other biometric methods, adversaries can also attack brainwave authentication by compromising different parts of the system [8]. The most applicable attack vector that targets specific users is arguably the *replay attack*, where the adversary injects a previously recorded sample of the biometric. Furthermore, with the current advance of machine learning techniques, it is also possible to generate fake brainwave data using Generative Adversarial Networks [59]. In this regard, if the authentication stimuli vary for each authentication attempt (order, type), the elicited brain responses will vary accordingly, but still provide the required user-specific features. This type of challenge-response protocol, implies that the attacker should be able to output results interactively in real-time, as the stimuli are not known in advance, which makes the attack harder to implement. Furthermore, an attacker observing a user while authenticating learns nothing about the brainwaves. Mimicry attacks, which are feasible for other biometrics (voice, gait), are not applicable because the adversary can not imitate non-volitional user responses.

The acquisition of EEG signals also raises privacy issues because brainwaves correlate e.g., with our mental states, cognitive abilities, and medical conditions [69]. An adversary that controls the authentication stimuli, such as an honest-but-curious authentication provider, could manipulate them to infer private data. Martinovic *et al.* [44] demonstrated the feasibility of this type of attacks. They successfully proved that, by manipulating visual stimuli, EEG signals could reveal users' private information about their bank cards, PIN numbers, area of living, and if they knew a particular person. Frank *et al.* [27] go even further, showing that it is possible to extract private data from EEG recordings using subliminal stimuli (short duration images embedded in visual content) that cannot even be consciously detected by users.

With the potential wide adoption of BCI applications in our everyday lives, security and privacy concerns are rising [9,13]. Our user study and other previous research [45] show that users are concerned about 'mind reading', but some people are already giving their brainwaves to third parties that offer brain-controlled games or relaxation applications. It is therefore paramount to research the security and privacy implications of using brainwaves in computer systems and work to design appropriate countermeasures before mainstream adoption.

7.3 Practical Implementation Aspects

Time to authenticate. A prototype implementation based on the N400:Faces brainwave authentication algorithm would require an initial *enrollment phase*. This means approximately 1 minute of brain data recording while the user looks at images in their PC. This phase could be extended to collect a higher amount of samples for training the system and broken into several shorter sessions for user convenience. It would be useful to implement a sample quality detector to adapt the duration of the enrollment process, similar to how fingerprint systems ask the user to place the finger in different angles until enough data is gathered for successful operation. Next, the *authentication phase* would require a minimum of 6 seconds to authenticate the user, though this time will vary due to the FRR. Fallback mechanisms should be implemented in case the authentication does not succeed in a reasonable time. Based on previous empirical research [75], the average time to authenticate with 8-character random passwords is around 7.5 seconds (12.8-13.2 seconds in tablet/smartphones [75]). Therefore, brainwave authentication is better in a best-case execution. But even if it takes longer, it has to be considered that usability perceptions can deviate from objective performance measures. For example, research shows evidence that graphical authentication schemes are perceived as more joyful than passwords even if the login time may exceed that of passwords [42,76]. In this sense, the N400:Faces is promising given the positive ratings on enjoyability obtained in the user study.

Extended Comparison. We use the framework of Bonneau *et al.* [14] to compare brainwave authentication against passwords (the most common solution) and fingerprint (the most used biometric). Table 5 summarizes this comparison according to the 25 criteria provided by the framework, grouped in *usability*, *deployability*, and *security* benefits. It can be seen that brainwave authentication provides better usability than passwords, and it could be comparable to that of fingerprints when FRR improves. On the security criteria, brainwaves bring additional benefits because they are not observable and can not be mimicked. Targeted impersonation attacks with synthetic or replayed data can be countered using the challenge(stimulus)-response nature of the brainwave authentication protocol. This allows the system to check response freshness and whether reactions correspond to stimuli that are meaningful for the legitimate user. Furthermore, as the adversary would need to interact with a legitimate authentication provider to obtain those per-user stimuli, we get resilience to phishing. The main security challenge is to reduce the FAR. Besides, brainwaves have the worst deployability according to the framework criteria, though these criteria focus on applicability to web authentication. Aspects like browser compatibility could be addressed by implementing brainwave authentication as part of the FIDO/WebAuthn protocols [77], currently supported in modern browsers. Additionally, there are other domains and use-cases outside the web realm where brainwaves could become practical.

Use-cases. The proposed brainwave authentication system was initially conceptualized for accessing PC applications, but the visual stimuli can be easily adapted to other devices and scenarios. Furthermore, once authenticated with the brainwave protocol, the user continues to have measurable brain activity, which can be leveraged for continuous authentication while wearing the headset. Brainwaves can be practical when users already wear an EEG reader for another application and a keyboard is inconvenient/unavailable. For example, authentication in Virtual Reality (VR) applications is still challenging as passwords are clearly unpractical. But modern VR headsets are introducing EEG sensors, making them a perfect scenario to apply our mechanisms. Additionally, with the ongoing miniaturization and integration of EEG sensors in devices that people commonly use (e.g., earbuds), having to carry them can be less problematic . Moreover, brainwaves could be augmented with other sensors that collect implicit biometrics (e.g., eye gaze) to improve authentication accuracy and, therefore, increase security.

7.4 Limitations

We acquired brainwaves in a lab environment and during a single recording session but we do not evaluate reliability and robustness with regard to noise or changing conditions. Nevertheless, based on previous research, we expect our system

	Usability								Deployability						Security										
Scheme	Memorywise-Effortless	Scalable-for-Users	Nothing-to-Carry	Physically Effortless	Easy-to-Learn	Efficient-to-Use	Infrequent-Errors	Easy-Recovery-from-Loss	Accessible	Negligible-Cost-per-User	Server-Compatible	Browser-Compatible	Mature	Non-Proprietary	Resilient-to-Observation	Resilient-to-Targeted-Impersonation	Resilient-to-Throttled-Guessing	Resilient-to-Unthrottled-Guessing	Resilient-to-Internal-Observation	Resilient-to-Leaks-from-Other-Verifiers	Resilient-to-Phishing	Resilient-to-Theft	No-Trusted-Third-Party	Requiring-Explicit-Consent	Unlinkable
Passwords			●		●	●	○	●	●	●	●	●	●	●		○						●	●	●	●
Fingerprint	●	●	●	○	●	●	●		○				●		○		●						●	●	
Brainwaves	●	●		●	●	○		●	○						●	●	●		●	●	●	●	●	●	

Table 5: Comparison of N400:Faces brainwave authentication against passwords and fingerprint using the framework by Bonneau et al. [14]. We use "•" to indicate that the scheme provides the benefit; and "◦"to denote that the benefit is somewhat provided.

to be robust as ERPs are less sensitive to background noise than continuous EEGs and, even if latency/amplitude might vary with external factors like stress, tiredness, etc. [19], ERPs reflect morphological components (e.g., skull thickness) that are more stable [5, 63]. Additional experiments in real-life conditions should be conducted to validate this hypothesis. In our experiments, we observed a high variability in the performance of different brainwave authentication tasks. We speculate that the number and quality of registered samples impacts the results, but further research is required to understand the factors inducing this variability and how to reduce their effect. It would be also valuable to investigate the scalability of the results to even larger populations.

With regard to usability, our user study is based on a sample of the population that includes generally young and technically-savvy users Bigger and diverse sets of users would yield a more comprehensive picture of the usability issues in brainwave authentication systems. We described the system to our participants embedding it in a realistic use case: we told them that they would have to watch one task out of the set of tasks in the experiment once a day, and this would replace the need to type passwords for their applications. With this description, a perfect implementation is assumed. The main methodological limitation is that we rely on self-reported qualitative feedback about intended future behavior based on participants perception of the described system, which might not accurately reflect reality [39]. With these characteristics, our goal is to describe problems that could hinder the adoption of brainwave-based authentication to consider when designing actual prototypes or experiments, but we do not claim any generalizable findings. Nonetheless, to achieve ecological validity, we need to test and evaluate the actual usability of authentication prototypes in real scenarios.

8 Conclusion

We contribute to the literature on behavioral biometrics with the first comparative study on the usability and performance of brainwave authentication protocols based on endogenous Event Related Potentials using consumer-grade EEG readers. Our results show the feasibility of authentication by recording brain activity while users look to short sequences of visual stimuli (images or words). With regard to perceived usability, users are positive about this type of systems but call for simpler headsets and fast authentication times. Considering participants feedback, we highlight the need to conduct extensive privacy research before brainwave-based applications become mainstream. When contextualizing our results, we found out that comparability with other works is hampered by differences in experimental conditions and performance reporting schemes, but also because the sample sizes used in the literature are very small (the majority ≤ 10). We therefore contribute our dataset to improve the availability of samples and provide a source for common benchmarking. To bridge the comparability gap, the authentication community should strive to establish a consistent approach for communicating performance metrics.

Acknowledgments

We thank our shepherd Deepak Kumar, as well as the anonymous reviewers who helped to improve this paper with their useful feedback. The work of P. Arias-Cabarcos has been supported through an Alexander von Humboldt Post-Doctoral Fellowship. This work was also funded by the Helmholtz Association (HGF) through the Competence Center for Applied Security Technology (KASTEL), topic "46.23 Engineering Secure Systems".

References

[1] Sherif Nagib Abbas, Mohammed Abo-Zahhad, and Sabah Mohammed Ahmed. State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. *IET Biometrics*, 4(3):179–190, September 2015.

[2] M. Abo-Zahhad, Sabah M. Ahmed, and Sherif N. Abbas. A new multi-level approach to EEG based human authentication using eye blinking. *Pattern Recognition Letters*, 82:216–225, 2016.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.

[4] Abdulaziz Almehmadi and Khalil El-Khatib. The state of the art in electroencephalogram and access control. In *2013 Third International Conference on Communications and Information Technology (ICCIT)*, pages 49–54, Beirut, Lebanon, June 2013.

[5] Blair C. Armstrong, Maria V. Ruiz-Blondet, Negin Khalifian, Kenneth J. Kurtz, Zhanpeng Jin, and Sarah Laszlo. Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. *Neurocomputing*, 166:59–67, 2015.

[6] Corey Ashby, Amit Bhatia, Francesco Tenore, and Jacob Vogelstein. Low-cost electroencephalogram (EEG) based authentication. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pages 442–445, May 2011.

[7] Michael P Barham, Gillian M Clark, Melissa J Hayden, Peter G Enticott, Russell Conduit, and Jarrad AG Lum. Acquiring research-grade erps on a shoestring budget: A comparison of a modified emotiv and commercial synamps eeg system. *Psychophysiology*, 54(9):1393–1404, 2017.

[8] Karen Becker, Patricia Arias-Cabarcos, Thilo Habrich, and Christian Becker. Poster: Towards a framework for assessing vulnerabilities of brainwave authentication systems. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 2577–2579, 2019.

[9] Sergio López Bernal, Alberto Huertas Celdrán, Gregorio Martínez Pérez, Michael Taynnan Barros, and Sasitharan Balasubramaniam. Cybersecurity in brain-computer interfaces: State-of-the-art, opportunities, and future challenges. *arXiv preprint arXiv:1908.03536*, 2019.

[10] Niels Birbaumer and Leonardo G Cohen. Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of physiology*, 579(3):621–636, 2007.

[11] D. H. R. Blackwood and W. J. Muir. Cognitive brain potentials and their application. *British Journal of Psychiatry*, 157(S9):96–101, December 1990.

[12] Maria V Ruiz Blondet, Sarah Laszlo, and Zhanpeng Jin. Assessment of permanence of non-volitional eeg brainwaves as a biometric. In *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, pages 1–6. IEEE, 2015.

[13] Tamara Bonaci, Ryan Calo, and Howard Jay Chizeck. App stores for the brain: Privacy & security in brain-computer interfaces. In *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*, pages 1–7. IEEE, 2014.

[14] Joseph Bonneau, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Proc. IEEE Symp. on Security and Privacy*, pages 553–567, 2012.

[15] Katharine Brigham and B. V. K. Vijaya Kumar. Subject identification from electroencephalogram (EEG) signals during imagined speech. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, Washington, DC, September 2010.

[16] John Brooke, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L McClelland. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[17] Ulrich Burgbacher and Klaus Hinrichs. An implicit author verification system for text messages based on gesture typing biometrics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2951–2954, 2014.

[18] Ann Cavoukian and Alex Stoianov. *Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy*. Information and Privacy Commissioner, Ontario, 2007.

[19] Hui-Ling Chan, Po-Chih Kuo, Chia-Yi Cheng, and Yong-Sheng Chen. Challenges and future perspectives on electroencephalogram-based biometrics in person recognition. *Frontiers in neuroinformatics*, 12:66, 2018.

[20] John Chuang, Hamilton Nguyen, Charles Wang, and Benjamin Johnson. I think, therefore I am: Usability and security of authentication using brainwaves. *Lecture Notes in Computer Science*, 7862 LNCS:1–16, 2013.

[21] Rig Das, Emanuele Maiorana, and Patrizio Campisi. Eeg biometrics using visual stimuli: A longitudinal study. *IEEE Signal Processing Letters*, 23(3):341–345, 2016.

[22] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. Now you see me, now you don't: protecting smartphone authentication from shoulder surfers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2937–2946, 2014.

[23] Jacques B. Debruille, Jaime Pineda, and Bernard Renault. N400-like potentials elicited by faces and knowledge inhibition. *Cognitive Brain Research*, 4(2):133–144, 1996.

[24] Matthieu Duvinage, Thierry Castermans, Mathieu Petieau, Thomas Hoellinger, Guy Cheron, and Thierry Dutoit. Performance of the emotiv epoc headset for p300-based applications. *Biomedical engineering online*, 12(1):56, 2013.

[25] Emotiv Systems. Emotiv EEG Headset Comparison Page, Url: https://www.emotiv.com/comparison/, Accessed: 31.07.2019.

[26] EU. General Data Protection Regulation, URL: https://gdpr-info.eu/, Accessed: 30.04.2019.

[27] Mario Frank, Tiffany Hwu, Sakshi Jain, Robert Knight, Ivan Martinovic, Prateek Mittal, Daniele Perito, and Dawn Song. Subliminal Probing for Private Information via EEG-Based BCI Devices. *CoRR*, abs/1312.6052:1–12, December 2013.

[28] Qiong Gui, Maria V. Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. A Survey on Brain Biometrics. *ACM Computing Surveys*, 51(6):1–38, 2019.

[29] Lindsay F Haas. Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 2003.

[30] InteraXon Inc. Url: https://choosemuse.com/, Accessed: 05.02.2020.

[31] ISO ISO. Iec 19795-1: Information technology–biometric performance testing and reporting-part 1: Principles and framework. *ISO/IEC, Editor*, 1(3):5, 2006.

[32] Isuru Jayarathne, Michael Cohen, and Senaka Amarakeerthi. Survey of EEG-based biometric authentication. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pages 324–329, November 2017.

[33] Preben Kidmose, David Looney, Lars Jochumsen, and Danilo P Mandic. Ear-eeg from generic earpieces: A feasibility study. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 543–546. IEEE, 2013.

[34] Els J Kindt. *Privacy and data protection issues of biometric applications*, volume 1. Springer, 2016.

[35] Belal Korany, Chitra R Karanam, Hong Cai, and Yasamin Mostofi. Xmodal-id: Using wifi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2019.

[36] Marta Kutas and Kara D. Federmeier. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual review of psychology*, 62(1):621–647, 2011.

[37] Marta Kutas and Steven A. Hillyard. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, 1980.

[38] Marta Kutas and Steven A. Hillyard. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163, 1984.

[39] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[40] Fabien Lotte, L Bougrain, A Cichocki, M Clerc, Marco Congedo, A Rakotomamonjy, and F Yger. A review of classification algorithms for EEG-based brain computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, 2018.

[41] Myndplay Ltd. Url: www.myndplay.com/, Accessed: 05.02.2020.

[42] Yao Ma and Jinjuan Feng. Evaluating usability of three authentication methods in web-based application. In *2011 Ninth International Conference on Software Engineering Research, Management and Applications*, pages 81–88. IEEE, 2011.

[43] Anthony J Mansfield and James L Wayman. Best practices in testing and reporting performance of biometric devices. 2002.

[44] Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito, Tomas Ros, and Dawn Song. On the feasibility of side-channel attacks with brain-computer interfaces. In *Proceedings of the 21st USENIX conference on Security symposium*, pages 34–43, Bellevue, WA, USA, August 2012.

[45] Nick Merrill, Max T. Curran, and John Chuang. Is the future of authenticity all in our heads? moving passthoughts from the lab to the world. In *Proceedings of the 2017 New Security Paradigms Workshop*, NSPW 2017, page 70–79, New York, NY, USA, 2017. Association for Computing Machinery.

[46] Matthew B Miles and A Michael Huberman. *Qualitative data analysis: An expanded sourcebook.* sage, 1994.

[47] Chisei Miyamoto, Sadanao Baba, and Isao Nakanishi. Biometric person authentication using new spectral features of electroencephalogram (EEG). In *2008 International Symposium on Intelligent Signal Processing and Communications Systems*, pages 1–4, February 2009.

[48] Kusuma Mohanchandra. Using Brain Waves as New Biometric Feature for Authenticating a Computer User in Real-Time. *International Journal of Biometric and Bioinformatics*, 7(1):49–57, 2013.

[49] Isao Nakanishi, Sadanao Baba, and Shigang Li. Evaluation of Brain Waves as Biometrics for Driver Authentication Using Simplified Driving Simulator. In *2011 International Conference on Biometrics and Kansei Engineering*, pages 71–76, Takamatsu, Japan, September 2011.

[50] Isao Nakanishi, Sadanao Baba, and Chisei Miyamoto. EEG based biometric authentication using new spectral features. In *2009 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 651–654, Kanazawa, Japan, December 2009.

[51] Isao Nakanishi and Masashi Hattori. Biometric potential of brain waves evoked by invisible visual stimulation. In *2017 International Conference on Biometrics and Kansei Engineering (ICBAKE)*, pages 94–99. IEEE, 2017.

[52] Isao Nakanishi and Takehiro Maruoka. Biometric authentication using evoked potentials stimulated by personal ultrasound. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, pages 365–368. IEEE, 2019.

[53] Isao Nakanishi and Takuya Yoshikawa. Brain waves as unconscious biometrics towards continuous authentication - the effects of introducing PCA into feature extraction. In *2015 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 422–425, Nusa Dua, Indonesia, November 2015.

[54] NeuroSky. NeuroSky MindWave Family Description Page, URL: http://neurosky.com/about-neurosky/, Accessed: 30.04.2019.

[55] R Palaniappan and D P Mandic. Biometrics from Brain Electrical Activity: A Machine Learning Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):738–742, 2007.

[56] Ramaswamy Palaniappan. Multiple Mental Thought Parametric Classification: A New Approach for Individual Identification. *International Journal of Signal Processing*, 2(4):222–226, 2005.

[57] Jeunese Payne, Graeme Jenkinson, Frank Stajano, M Angela Sasse, and Max Spencer. Responsibility and tangible security: Towards a theory of user acceptance of security tokens. *arXiv preprint arXiv:1605.03478*, 2016.

[58] Jonathan Peirce, Jeremy R Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, 2019.

[59] Tanya Piplani, Nick Merill, and John Chuang. Faking it, making it: Fooling and improving brain-based authentication with generative adversarial networks. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2018.

[60] Marios. Poulos, Maria Rangoussi, and Nikolaos Alexandris. Neural network based person identification using EEG features. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, volume 2, pages 1117–1120, March 1999.

[61] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *Journal of machine learning research*, 5(Jan):101–141, 2004.

[62] Maria V Ruiz-Blondet, Zhanpeng Jin, and Sarah Laszlo. Cerebre: A novel method for very high accuracy event-related potential biometric identification. *IEEE Transactions on Information Forensics and Security*, 11(7):1618–1629, 2016.

[63] Maria V. Ruiz Blondet, Sarah Laszlo, and Zhanpeng Jin. Assessment of permanence of non-volitional EEG brainwaves as a biometric. In *IEEE International Conference on Identity, Security and Behavior Analysis*, pages 1–6, Hong Kong, China, March 2015.

[64] Scott Ruoti, Brent Roberts, and Kent Seamons. Authentication melee: A usability analysis of seven web authentication systems. In *Proceedings of the 24th International Conference on World Wide Web*, pages 916–926, 2015.

[65] Phattarapong Sawangjai, Supanida Hompoonsup, Pitshaporn Leelaarporn, Supavit Kongwudhikunakorn, and

Theerawit Wilaiprasitporn. Consumer grade eeg measuring sensors as research tools: A review. *IEEE Sensors Journal*, 2019.

[66] Javad Sohankar, Koosha Sadeghi, Ayan Banerjee, and Sandeep K.S. Gupta. E-BIAS:A Pervasive EEG-Based Identification and Authentication System. In *Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pages 165–172, Cancun, Mexico, November 2015.

[67] Nancy K Squires, Kenneth C Squires, and Steven A Hillyard. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4):387–401, 1975.

[68] Shridatt Sugrim, Can Liu, Meghan McLean, and Janne Lindqvist. Robust Performance Metrics for Authentication Systems. In *Proceedings 2019 Network and Distributed System Security Symposium*, Reston, VA, February 2019. Internet Society.

[69] Shravani Sur and VK Sinha. Event-related potential: An overview. *Industrial Psychiatry Journal*, 18(1):70, 2009.

[70] Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue. A survey of keystroke dynamics biometrics. *The Scientific World Journal*, 2013, 2013.

[71] Kavitha P Thomas, AP Vinod, et al. Eeg-based biometrie authentication using self-referential visual stimuli. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3048–3053. IEEE, 2017.

[72] Julie Thorpe, Paul C. van Oorschot, and Anil Somayaji. Pass-thoughts: authenticating with our minds. In *Proceedings of the 2005 workshop on New security paradigms - NSPW '05*, pages 45–56, Lake Arrowhead, CA, USA, September 2005.

[73] Marijn van Vliet, Christian Mühl, Boris Reuderink, and Mannes Poel. Guessing What's on Your Mind: Using the N400 in Brain Computer Interfaces. In *Lecture Notes in Computer Science*, volume 6334 LNAI, pages 180–191. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[74] Marijn van Vliet, Arne Robben, Nikolay Chumerin, Nikolay V Manyakov, Adrien Combaz, and Marc M Van Hulle. Designing a brain-computer interface controlled video-game using consumer grade eeg hardware. In *2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, pages 1–6. IEEE, 2012.

[75] Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. Honey, i shrunk the keys: Influences of mobile devices on password composition and authentication performance. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, NordiCHI '14, page 461–470, New York, NY, USA, 2014. Association for Computing Machinery.

[76] Emanuel Von Zezschwitz, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 277–286, 2013.

[77] W3C. Web Authentication: An API for accessing Public Key Credentials Level 2. W3C Candidate Recommendation Snapshot, 2020.

[78] Frederick W Wheeler, Richard L Weiss, and Peter H Tu. Face recognition at a distance system for surveillance applications. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2010.

[79] Jonathan Wolpaw and Elizabeth Winter Wolpaw. *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, January 2012.

[80] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767 – 791, 2002.

[81] Mahendra Yadava, Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. Analysis of eeg signals and its application to neuromarketing. *Multimedia Tools and Applications*, 76(18):19087–19111, 2017.

[82] Su Yang and Farzin Deravi. On the Usability of Electroencephalographic Signals for Biometric Recognition: A Survey. *IEEE Transactions on Human-Machine Systems*, 47(6):958–969, 2017.

[83] Hui-yen Yap, Yun-huoy Choo, and Wee-how Khoh. Overview of Acquisition Protocol in EEG Based Recognition System. In Zeng et al., editor, *Brain Informatics*, volume 10654, pages 129–138. Springer, Cham, Switzerland, 2017.

Appendix: Open Data

The anonymized dataset and experiment material (script, questionnaire, codebooks) are available at `https://git.scc.kit.edu/kr2925/brainwave-authentication`.