

# Classifying Mental Gestures with In-Ear EEG

Nick Merrill<sup>1</sup>, Max T. Curran<sup>1</sup>, Jong-Kai Yang<sup>1</sup>, and John Chuang<sup>1</sup>

<sup>1</sup>BioSENSE Lab, School of Information, UC Berkeley

{nick, mtcurran, jong-kai.yang, chuang}@ischool.berkeley.edu

**Abstract**—While brain-computer interfaces (BCI) based on electroencephalography (EEG) have improved dramatically over the past five years, their inconvenient, head-worn form factor has challenged their wider adoption. In this paper, we investigate how EEG signals collected from the ear could be used for "gestural" control of a brain-computer interface (BCI). Specifically, we investigate the efficacy of a support vector classifier (SVC) in distinguishing between mental tasks, or gestures, recorded by a modified, consumer headset. We find that an SVC reaches acceptable BCI accuracy for nine of the subjects in our pool (n=12), and distinguishes at least one pair of gestures better than chance for all subjects. User surveys highlight the need for longer-term research on user attitudes toward in-ear EEG devices, for discreet, non-invasive BCIs.

## I. INTRODUCTION

Brain-computer interfaces (BCIs) enable the control of a computer without muscular movement. BCIs based on electroencephalography (EEG) are popular due to their use of non-invasive surface electrodes.

The hardware that drives EEG-based BCIs has improved dramatically over the past five years, decreasing in size and cost by orders of magnitude [1]. Many consumer devices leverage this technology: as of February 2016, there are at least five EEG devices on the market, ranging from 100 to 500 USD, and featuring one to sixteen electrodes. Many of them transmit data wirelessly to computers and smart devices. Meanwhile, advances in machine learning have radically improved the reliability of BCI applications. Taken together, prospects seem bright for the wider adoption of BCIs in everyday life.

However, the head-worn form-factor, and awkward visibility of EEG-based BCIs has proven a stubborn challenge to BCI adoption [2]. Both disabled and healthy subjects complain about the comfort of head-worn devices, the difficulty of applying electrodes correctly to the scalp, and questionable aesthetics of wearing such a visible device in public, social settings [3], [4].

One possible solution to this problem is to embed EEG electrodes in earbuds, collecting EEG signals from the ear canal. While early work framed in-ear EEG largely as a tradeoff between ergonomics and signal quality [5], in-ear EEG signals are at least robust enough to detect auditory evoked responses [6], more recent work has indicated that EEG collected in the ear may have its own, unique affordances. For example, one study built a rudimentary eye-tracker using muscular signals (EMG, or electromyography) collected from the ear canal [7].

The performance of *mental gestures* has proven a popular paradigm in past research on scalp-based BCIs [8], [9]. Until

now, no studies have explored mental-gesture based BCIs using in-ear EEG. Which gestures are well-suited to signals collected from the ear, and how people rate the usability of gesture-based in-ear EEG, remains largely a mystery.

In this paper, we take an exploratory step toward gesture-based BCI from EEG data collected at the ear. We collect a corpus of mental gestures from a number of subjects, and, for each subject, we simulate calibrating a binary BCI, and estimate its accuracy. We also collect self-report data from subjects about the usability of tasks, and of the device.

We find that we can calibrate a decent binary BCI for most subjects. In line with work for traditional, EEG-based BCI, we find that calibrated BCI accuracy correlates strongly to the average classifiability of gesture for that study, evidence of BCI "literacy" or competence. We do not find evidence that different gestures lead to better or worse classifiability. However, we do find that subjects rate some gestures as easier and more engaging to perform and repeat. Finally, we find that subjects are divided on attitudes toward wearing an EEG device in their ear. Our findings highlight the need for longer-term usability probes to gauge the usefulness and desirability of in-ear EEG in everyday life.

## II. BACKGROUND

Generally, BCI systems aim to recognize a user's mental gestures as one of a finite set of discrete symbols, a problem that can be framed as a pattern recognition task [10]. The difficulty of this task stems primarily from the fact that symbols are expressed differently between individuals (they are *idiosyncratic*) [11].

### A. Adaptive Calibration and Classification

In order to compensate for variability in BCI signals, recent work has leveraged adaptive classification algorithms to distinguish between *mental gestures* [10], [11], [8]. Automated calibration procedures have turned BCI novices into competent users over the course of hours instead of days or weeks, and without manual calibration [11]. During calibration, users perform *labeled* (i.e. known) mental gestures, in order to produce training data for a classifier.

Relevant to our study are support vector classifiers (SVCs), which create models that can predict categorical classes of unlabeled data. Past work has found particular success using linear SVCs in BCI application, as opposed to SVCs with non-linear kernels, or other non-linear classifiers such as neural networks [12], [1]. Linear SVCs select the linear hyperplane that maximizes distance from the nearest training points, which

increases the model’s generalizability. The effectiveness of linear SVCs compared to other classifiers have led some to believe that classification in EEG-based BCI is fundamentally linear [10].

### B. Classifiability of Gestures Collected from the Ear

Due to linear support vector classifiers’ common use in gesture-based BCIs, the ability of an SVC to distinguish between gestures is a useful way to gauge the potential for developing an effective BCI based around the performance of mental gestures. [8], [13] In this paper, we collect a corpus of mental gestures, generate all possible pairs of gestures, and estimate each pair’s distinguishability to the classifier.

The results of this procedure can shed light on two major issues when developing a novel BCI. First, we can develop a rough, best-case estimate of our BCI’s reliability, as our gestures are recorded in a controlled setting, and we do not account for changes in gesture expression over time. Second, by seeing which tasks appear most frequently in best-case gesture pairs, this procedure can help us form hypotheses about which gestures to include in standard calibration routines.

Finally, this procedure allows us to hypothesize about which sources of signal are most informative. If we see certain types of gestures frequently appearing in best-case gesture pairs, we can begin to generate hypotheses about what sorts of signal in-ear EEG is well-suited to pick up.

## III. SELECTING MENTAL GESTURES

TABLE I  
LIST OF MENTAL GESTURES.

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

We primarily chose tasks that showed promise in previous experiments [9], [13], though generally we found little research to suggest which mental tasks result in the most robust EEG signals. We attempted to include a variety of tasks to draw on different EEG signals. The breath, song, listen, and face tasks were performed with the participants’ eyes closed, while the cube task was performed with the participants’ eyes open. In addition, the song and face tasks involved an additional unique choice by the participant.

## IV. DATA COLLECTION

12 graduate students (7 male, 5 female, with a mean age of  $28 \pm 4.36$ ) from a large university on the west coast of the United States completed our study protocol, which was approved by the the university’s Committee for Protection of Human Subjects (CPHS). Study procedures began with an informed consent process, followed by a demographics questionnaire, a set up period with the EEG device, completion

of a set of 5 tasks presented on a laptop while EEG was recorded, and finally a post-experiment questionnaire. We used a Neurosky Mindwave Mobile wireless EEG headset, which is sold online to the general public for 99.99 USD. Modifications made to the original included releasing the sensing electrode from the plastic forehead arm, removing the electrode, and replacing it by soldering a new 6 mm gold cup electrode onto the original wire. The gold cup electrode was bent to allow for a comfortable fit in a range of ear canal sizes. The device being worn is shown in Figure 1.



Fig. 1. Modified Neurosky Mindwave setup.

To set up the device, the experimenter placed the sensing electrode in the ear canal against the superior wall (facing upward) with a rolled foam earplug placed beneath it to keep the electrode comfortably in place. The Mindwave device transmits data wirelessly via Bluetooth, so it was paired and the connection was confirmed before beginning the task phase of the experiment.

Table I lists the tasks performed by participants. Instructions for tasks were presented visually using PsychoPy [14] and read aloud verbally by the experimenter. We asked participants to sit in a comfortable position and remain as still as possible for all tasks. Participants used a wireless remote held comfortably in their laps to begin each task when they were ready. Each task was recorded during two sets of 5 trials each to lessen boredom effects and each trial was 12 seconds in length.

## V. DATA PROCESSING

### A. Producing Feature Vectors

Following [13], we use logarithmic binning to produce compressed feature vectors of a variable size. This technique has been show to offer robust, linear classifiability in healthy subjects. It is unique in its use of the entire frequency spectrum. Since EEG activity is associated with frequencies from 1-40Hz, we presume this range contains the majority of relevant signal. However, we do not rule out the possibility that useful signal exists in other frequency ranges. Muscular activity, for example, might be correlated with mental gestures in some cases. Logarithmic binning produces feature vectors biased toward known sources of signal, while still including data points from outside this frequency range that may be informative.

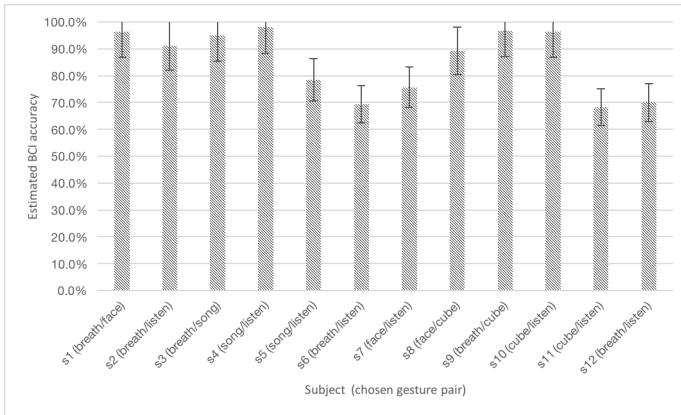


Fig. 2. BCI performance for each participant, estimated by classifier accuracy on best-case gesture pair.

### B. Linear SVC

We analyzed the EEG signals collected during the tasks using a support vector classifier (SVC). Since past work has shown that classification tasks in EEG-based BCI are linear [10], we used LIBLINEAR, [15], a popular linear SVC kernel. For each task, for each participant, 120 seconds of data was collected in total across 10 trials of 12 seconds each. We initially tried analyzing all 12 seconds of data per trial, but found that removing the first 2 seconds from the beginning, and last second from the end of each trial to account for the transition to and from performing a given task improved our results. Following preprocessing (see Producing Feature Vectors), we have 30 samples per participant, per task.

### C. Simulating the Calibration of a Binary BCI

Of the mental gestures in the dataset, we seek to identify, for each subject, the pair of gestures that produces the highest accuracy with our Linear SVC. This will result in a personalized, binary (two-class) classifier, where the SVC can discriminate between two mental gestures performed by the subject with the highest classification accuracy. The gesture-pairs may vary from subject to subject. For example, one subject’s best-case pair may be *breath* and *song*, while another’s may be *cube* and *face*.

To simulate calibrating a binary BCI for a given subject, we generate every possible pair of gestures. For each gesture, we perform seven-fold cross-validation, whereby the gesture data in question are split into different sets of training and testing data. This allows us to estimate the performance of our classifier on unknown samples, assuming samples in the future are drawn from the same distribution as the samples in our training set. The output of this process is, for each pair of gestures, a mean accuracy, and standard deviation between accuracies over the seven folds of cross-validation.

## VI. SIMULATED BCI RESULTS

We simulated the calibration of a binary BCI by training and testing an SVC for all possible pairs of gestures, for each participant, and choosing each participant’s highest-scoring

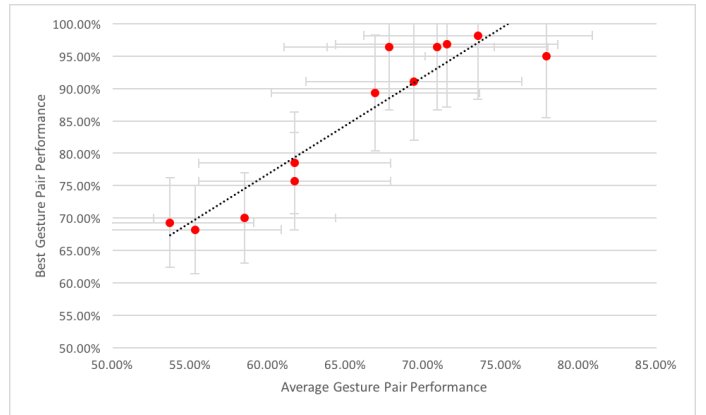


Fig. 3. For each subject, best-case BCI performance (the estimated accuracy of the top-scoring pair of gestures), against the average gesture pair performance for that participant.

(most accurate) gesture pair. Estimated BCI accuracy across all subjects was an average of 85.4% ( $\sigma = 12.1\%$ ) (Figure 2). Nine of the twelve subjects in our pool reached Vidaurre and Blankertz’s (2010) threshold of BCI literacy (75%) [16] Six of twelve subjects achieved estimated accuracies of over 90%.

All gestures we gave to our users appeared in at least one best-case gesture pair (Table I). Listen and breath appeared 8 and 6 times, respectively, though our sample size was too small to establish statistical significance (24 observations, or two per participant). When we considered only gesture pairs that achieved over 75% threshold accuracy, we found that all gestures still appeared in at least one best-case pair, and gestures appeared with roughly the same frequency across best-case pairs.

Across all possible pairs of gestures, among all subjects, we find no correlation between gesture and classifier accuracy. We also do not find any correlation between the device’s reported signal quality and classifier accuracy. However, we find a strong correlation between a subject’s average classifier accuracy (on all gesture pairs), and a subject’s best-case accuracy ( $r = .9334$ ,  $p < 0.001$ ) (Fig 3). We do not find evidence that classifier accuracy correlates to any of our subject-reported data (task ease of use, repeatability, willingness to use again), nor to any of the demographic variables we collected.

## VII. USABILITY RESULTS

Figure 4 shows the mean responses for the post-experiment questionnaire’s 7-point Likert-type usability scales. The questionnaire asked participants to rate the tasks on three usability scales with labeled extremes: “Please rate each of the experimental tasks on ease of performing. (very difficult - very easy),” “Please rate each of the experimental tasks on how engaging/interesting they were to perform. (very boring - very engaging),” and “Please rate each of the experimental tasks on how easy they would be to repeat often. (very difficult to repeat - very easy to repeat).” The results of nine participants are shown here, as the three experimenters’ responses were not included.

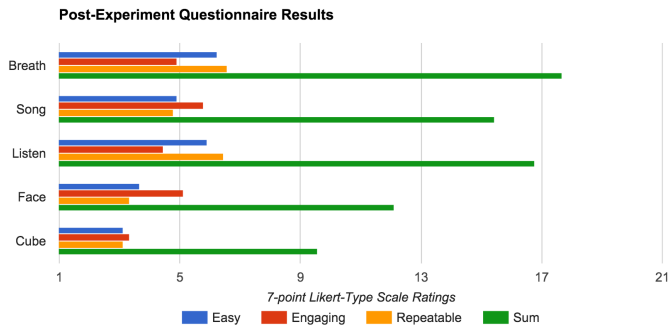


Fig. 4. Post-experiment questionnaire mean responses on a 7-point likert-type scale.

The song and face tasks included a choice by the participant of which song or face to imagine and the experimenter recorded these choices. For the face task, all participants except one chose the face of someone with whom they had a personal relationship like a family member, friend, or significant other; the participant who did not follow this trend chose a famous politician. One participant commented that the face task elicited an emotional response,

*"Imagining my grandmother led to many other nice associations, including compassion"*

Another optional, open response question asked, "Do you think you would be likely to wear an in-ear EEG device in a real life situation?" Three subjects provided negative reviews.

*"I currently can't imagine wanting to put something in my ear (even earbuds are not comfortable), but maybe, if it was really unobtrusive"*

*"Not particularly likely; only if everyone's doing it"*

*"Maybe as specialized work equipment"*

The remaining subjects reported that they might consider wearing in-ear EEG in their everyday life, framing their adoption as conditional on other factors.

*"Sure as long as it was easy to put on or wear"*

*"Sure! If I could do something interesting with it, depends on the application."*

*"Yes (if it gets smaller)"*

### VIII. LIMITATIONS

Our study included only twelve subjects, all of whom were fairly young (mean age of 28), and students at a university. Our study also tested a limited number of mental gestures. Future work should attempt to replicate our findings with a larger, more diverse sample.

We used a \$100 consumer device that we modified ourselves; it is unclear how robustly our device simulates the performance of a hypothetical consumer device that measures EEG at the ear. It is also unclear how the homemade appearance of the device affected participants' assessments of how likely they would be to wear an in-ear EEG device in the future.

Our questions on usability only covered participants' first experiences with the device, and with their gestures. From a technical standpoint, our study does not indicate if or how the performance of mental gestures may change over time. From a usability standpoint, our study also does not explore how attitudes toward usability (e.g. learning effects), nor attitudes toward the device, may change as subjects use in-ear EEG in daily life.

We collected data from participants while they were sitting down, and indoors. It is not clear how our results would generalize to ambulatory sensing environments (for example, how noise from muscular movement, or environmental electromagnetic noise, would effect classifier accuracy). Furthermore, participants were told that the data we were collecting was for a use-case of authentication. Although the questions and responses reported in Section VII were not specific to authentication, the use case may have affected users' responses, or the way users expressed their mental gestures.

Finally, the room in which our data were collected was not well-insulated from outside noises. Sounds of passing people, and hourly bells from the nearby bell tower, were clearly audible in our subject environment. It is unclear how these aural distractions affected our results, or if these distractions make our work more or less generalizable to performance in real-life settings.

### IX. DISCUSSION

This study investigated how well EEG signals, collected from the ear canal, could be used to calibrate a binary brain-computer interface, using mental gesture samples collected from twelve subjects. Generally, our subjects' estimated accuracy was good: average best-case accuracy was 85.4% ( $\sigma = 12.1\%$ ) (Figure 2), and nine of the twelve subjects in our pool reached a threshold for "BCI literacy" [16]. Six of twelve subjects achieved estimated accuracies of over 90%.

Interestingly, we found no correlation between gesture and classifier accuracy, nor any correlation between classifier accuracy and signal quality, or any of the demographic variables we collected. However, we found a strong, linear correlation between a subject's estimated BCI accuracy, and the average classifier accuracy across all gesture pairs for that subject. This finding highlights the highly idiosyncratic nature of BCI "(il)literacy," a phenomenon noted in scalp-based BCIs, whereby some subjects show a general aptitude for using BCIs, while others struggle to achieve acceptable performance. Factors that contribute to BCI literacy are an ongoing subject of research in traditional BCIs.

While all gestures were equally likely to result in strong BCI accuracy, subjects did not rate all tasks as equally usable. Subjects rated the "breath" and "listen" gestures as easiest to perform, and easiest to repeat. The "song" gesture was rated as most engaging, but slightly less usable than "breath" and "listen." BCIs that use in-ear EEG should weigh the usability of gestures against their accuracy in customized classifiers. While a given gesture may afford a subject good accuracy,

the user may not adopt the BCI if this gesture is not easy or pleasant to perform in everyday life.

Finally, subjects were highly divided on the question of how likely they would be to wear an in-ear EEG device in real-life situations. However, these judgments were based on first encounters with our device, which was not built with ergonomics in mind, and subjects made these judgments in the absence of interface feedback. A longer-term usability probe could begin to investigate in greater depth usability and attitudes toward in-ear EEG sensing.

## X. FUTURE WORK

Our work provides a starting point for understanding how in-ear EEG could be used to build a mental-gesture based BCI. However, much work remains. While in-ear sensors could provide usability benefits in everyday settings, our work does little to explore how these benefits would be realized. One clear priority for future work is to examine the accuracy of classifiers in ambulatory settings: outdoors, and in a wide variety of movement contexts (walking, running, biking, etc).

In our study, we found evidence of general BCI literacy/illiteracy in individual subjects; future work could explore how in-ear BCI literacy relates to literacy with traditional, scalp-based BCIs, along with the phenomenon of BCI literacy more generally. Future studies could also experiment with a more diverse range of mental gestures, and with a larger subject pool. Such work could investigate what kinds of tasks are well-suited to in-ear, versus scalp-based BCIs.

Crucially, future work should examine user attitudes toward the device, and toward gestures and their performance, over longer timescales, and in more naturalistic settings. A longer-term, *in vivo* study could investigate the effects of practice with a gesture, how users relate to their EEG device, and how people feel being monitored by EEG, in a variety of different contexts and environments. Longer-term studies could reveal challenges and opportunities to usability, and surface factors that will drive or hinder future adoption of in-ear BCIs.

## XI. CONCLUSION

Our results indicate that in-ear BCI applications using mental gestures may be feasible. We also find evidence that individual differences in BCI "literacy," a common finding from traditional BCI work, also hold for in-ear BCIs. While in-ear EEG offers potential advantages to usability over scalp-based EEG, longer-term, *in vivo* studies will be necessary to fully realize these benefits. Examining the use of in-ear, mental-gesture based BCIs in different contexts, and in different social settings, will shed light on the user preferences and attitudes that will drive or hinder adoption of such systems in the future.

## ACKNOWLEDGMENTS

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

## REFERENCES

- [1] M. Grierson and C. Kiefer, "Better Brain Interfacing for the Masses: Progress in Event-related Potential Detection Using Commercial Brain Computer Interfaces," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '11. New York, NY, USA: ACM, 2011, pp. 1681–1686. [Online]. Available: <http://doi.acm.org/10.1145/1979742.1979828>
- [2] V. Mihajlovic, B. Grundlehner, R. Vullers, and J. Penders, "Wearable, Wireless EEG Solutions in Daily Life Applications: What are we missing?" *IEEE Journal of Biomedical and Health Informatics*, vol. Early Access Online, 2014.
- [3] J. I. Ekandem, T. A. Davis, I. Alvarez, M. T. James, and J. E. Gilbert, "Evaluating the ergonomics of BCI devices for research and experimentation," *Ergonomics*, vol. 55, no. 5, pp. 592–598, May 2012. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00140139.2012.662527>
- [4] W. D. Hairston, K. W. Whitaker, A. J. Ries, J. M. Vettel, J. C. Bradford, S. E. Kerick, and K. McDowell, "Usability of four commercially-oriented EEG systems," *Journal of Neural Engineering*, vol. 11, no. 4, p. 046018, Aug. 2014. [Online]. Available: <http://iopscience.iop.org/1741-2552/11/4/046018>
- [5] P. Kidmose, D. Looney, L. Jochumsen, and D. P. Mandic, "Ear-EEG from generic earpieces: A feasibility study," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 543–546.
- [6] P. Kidmose, D. Looney, and D. P. Mandic, "Auditory evoked responses from Ear-EEG recordings," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 586–589.
- [7] H. Manabe, M. Fukumoto, and T. Yagi, "Conductive rubber electrodes for earphone-based eye gesture input interface," *Personal and Ubiquitous Computing*, vol. 19, no. 1, pp. 143–154, 2015.
- [8] E. V. C. Friedrich, C. Neuper, and R. Scherer, "Whatever Works: A Systematic User-Centered Training Protocol to Optimize Brain-Computer Interfacing Individually," *PLoS ONE*, vol. 8, no. 9, p. e76214, Sep. 2013. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0076214>
- [9] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I Think, Therefore I Am: Usability and Security of Authentication Using Brainwaves," in *Financial Cryptography and Data Security*, ser. Lecture Notes in Computer Science, A. A. Adams, M. Brenner, and M. Smith, Eds. Springer Berlin Heidelberg, Jan. 2013, no. 7862, pp. 1–16. [Online]. Available: [http://link.springer.com/chapter/10.1007/978-3-642-41320-9\\_1](http://link.springer.com/chapter/10.1007/978-3-642-41320-9_1)
- [10] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of neural engineering*, vol. 4, 2007.
- [11] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Machine-Learning-Based Coadaptive Calibration for Brain-Computer Interfaces," *Neural Computation*, vol. 23, no. 3, pp. 791–816, Mar. 2011. [Online]. Available: [http://www.mitpressjournals.org/doi/abs/10.1162/NECO\\_a\\_00089](http://www.mitpressjournals.org/doi/abs/10.1162/NECO_a_00089)
- [12] D. Garrett, D. Peterson, C. Anderson, and M. Thaut, "Comparison of linear, nonlinear, and feature selection methods for eeg signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141–144, Jun. 2003. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1214704>
- [13] N. Merrill, T. Maillart, B. Johnson, and J. Chuang, "Improving Physiological Signal Classification Using Logarithmic Quantization and a Progressive Calibration Technique," *Proceedings of the 2nd International Conference on Physiological Computing Systems (PhyCS)*, 2015.
- [14] J. W. Peirce, "PsychoPy—psychophysics software in Python," *Journal of neuroscience methods*, vol. 162, no. 1, pp. 8–13, 2007.
- [15] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A Library for Large Linear Classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1390681.1442794>
- [16] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain topography*, vol. 23, no. 2, pp. 194–198, 2010.