Evoked Response Potential Training on a Consumer EEG Headset

James Derek Jacoby and Melanie Tory
Department of Computer Science
University of Victoria
Victoria, Canada
derekja@uvic.ca, mtory@uvic.ca

James Tanaka
Department of Psychology
University of Victoria
Victoria, Canada
jtanaka@uvic.ca

***Abstract*—Neurofeedback paradigms are of increasing interest, both for brain-computer interfaces and for training purposes. In the current study we examine multi-session training in a P3 speller application which uses P300 evoked response potentials to identify the letter of a word a participant is attempting to spell. Because neurofeedback training is most effective over time, there is an advantage to using low-cost hardware that can be used in the home rather than expensive research-grade equipment. This study is performed on an Emotiv Epoc EEG headset and some of the problems and workarounds inherent in this are discussed. This headset is sensitive enough to detect evoked response potentials and evidence of a training effect is found. Motivational aspects of ERP training are discussed and an alternate P300 training paradigm is presented.**

Keywords-P300; P3 speller; training; BCI2000; EEGLab

I. Introduction

The P3 speller is a brain-computer interface (BCI) for text input. It takes advantage of the P300, strongest and easiest to reliably detect of the evoked response potentials (ERP). This is a positive deflection across a wide number of electrode locations at approximately 300 milliseconds after a stimuli of interest is presented[1]. The P3 speller has been used in many contexts as both an assistive device and a general study paradigm. It has been noted that training effects occur, but very little systematic research has been done on these effects[2]. Training is important because some proportion of study participants simply fail to show a strong enough P300 signal to allow them to use the P3 speller. Others show robust effects during one test session but in another session are unable to get acceptable performance. This inter-subject and inter-session variability, as well as the general speed of text entry, are some reasons that the P3 speller has fallen out of general favor in the assistive device community in favor of eye tracking solutions[3].

We present a systematic study of training effects on the P3 speller using a consumer-grade EEG device. The goals in showing ERP training performance are twofold – first, the P3 speller is an interesting input platform in some situations and learning how to use it more effectively is necessary if a P3-based BCI is to be useful. Equally important is that this sustained visual attention ERP paradigm appears to be correlated with attentional deficits in conditions such as ADHD[4]; since there exist non-neurofeedback based attentional training protocols for ADHD[5], it is therefore conceivable to develop a neurofeedback-based training protocol for such conditions. It is in relation to this second point that the current study is completed using a consumer EEG device, the Emotiv Epoc. The device is relatively inexpensive ($500), opening up possibilities that neurofeedback training could be performed in the home and in therapeutic settings where a research-grade EEG (costing tens of thousands of dollars) would simply be impractical. In this work, we show evidence of a training effect with repeated P3 speller use, and we demonstrate that such effects can be detected with consumer-grade EEG hardware. Furthermore, we propose a related card-game paradigm which has the ability to manipulate cognitive factors of the task to maintain long term interest while still remaining engaged in the attentional training.

978-1-4673-7788-1/15/$31.00 ©2015 IEEE

II. Background and Related Work

Evoked response potentials have been studied since the 1930’s, but it is the close tying of specific stimuli presentation to EEG response that marks the modern age of ERP. The P300 was the first signal examined in that way[6]. A comprehensive review in the early 1990’s summarized factors leading to a robust P300 response, including novelty, attention, and the absence of certain medical conditions[1]. It was also around this time that the P3 speller paradigm was introduced, as the first ERP-controlled brain computer interface (BCI) where users focus attention on flashing stimuli and use the P300 response for the system to identify the focus[7].

The P3 speller task is painfully boring. This limits its utility as a long term training mechanism. Some work has looked at using P300 signals as an input modality in computer gaming[8]. This approach seems like a great first start to developing stimulus materials and test methodologies that are usable for a long enough period to see a training effect develop.

Attentional and motivational factors strongly affect P300 strength and reliability[9] and this is likely a reason that marked decrease in P300 response is often seen in individuals with attentional deficits[4][10]. This reduced P300 response has even been suggested as a diagnostic criterion for ADHD, but the situation is more complex. Intriguingly, in adults with ADHD, adaptations and training strategies are postulated as being able to at least partially overcome these deficits[11].

Direct attentional training is an effective intervention in children with ADHD[5]. Although general training for attentional conditions has been studied, the specific question of EEG training has received surprisingly little coverage. One of the few papers on the topic is a short article attempting to train users to improve the P300 stability, although this successfully focused on improving phase locking factors rather than overall magnitude of the P300 peak[2].

An early study showing the use of consumer grade EEG equipment in ERP research was presented in 2011. It showed that it was possible to use the Neurosky Mindwave single sensor headset to detect P300 responses[12]. This headset has been judged a bit too limited by other researchers, and the Emotiv Epoc has received more study, with the general conclusion that while it generates reliable data that it suffers from long-term comfort issues[13], reduced signal quality compared to research grade instruments, and a lack of precision in headset fitting[14].

The current study demonstrates effective ERP results on the Emotiv Epoc and shows preliminary evidence of training P300 strength through repeated sessions of practice. This training effect on low-cost hardware opens the door to ERP neurofeedback protocols which are accessible to individuals outside the formal research lab.

III. Methods

A. Participants

Eight University of Victoria undergraduates participated in exchange for experimental credit in their psychology classes. 2 of the 8 subjects had to be removed due to rejection criteria, leaving 6 subjects in the study. Rejection criteria were determined in advance to exclude participants who were unable to achieve consistent enough performance to enable the classifier to be trained. Participants returned for two sessions. 5 of the 6 subjects were female. Selection criteria included an absence of attentional disorders and lack of experience with related EEG studies. Four additional subjects were run to trial an alternate card game task, described below.

B. Apparatus

The current studies were performed on an EEG platform which is relatively low-cost and generally intended for computer gaming. The primary use of the Emotiv Epoc is for capturing slow-wave frontal lobe EEG signals. The most common tasks are relaxation trainers, meditation trainers, and other similar applications. The Emotiv uses saline-coupled electrodes — this is a simpler system for casual use than a standard gel-coupled headset, but the conductivity is not as good, particularly in participants with thick hair. It is particularly sensitive to motion artifacts, although in its reversed position in the studies reported here, those artifacts are more pronounced from full head movement than they are from simple eye blinks or facial movements. This particular headset is also sensitive to the size of the head. Unlike a standard EEG cap which comes in multiple sizes, the Emotiv is available in only a single size. Some participants with smaller than average heads have a very difficult time achieving an adequate headset fit.

The 14 channel headset is normally worn with the majority of the sensors on the frontal area, but it was reversed in order to

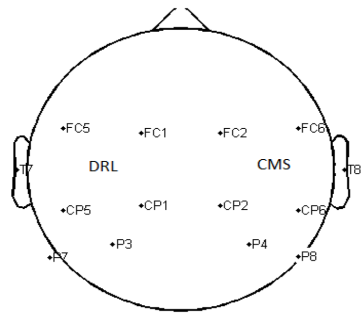

Figure 1: Approximate electrode positions, DRL is a noise cancellation electrode and CMS is the common reference

access the desired head locations. The electrodes were positioned as closely as possible to the locations described in figure 1, although head shape variability and a relatively non-adjustable headset did not allow these locations to be as precisely targeted as would have been the case with a cap-style electrode set. After positioning, the onscreen Emotiv control panel was used to ensure that all electrodes were shown as green in the user interface, indicating an impedance of 20K Ohm or less[15]. The rejection criterion for headset fitting was that no more than 2 of the electrodes could be at yellow, the rest at green. This attention to headset fitting reduced the rejection rate substantially from early piloting, but required up to 15 minutes of fitting time for some subjects. This careful attention ensured good connectivity and was re-checked in the middle of the session and whenever a participant took a break. If connectivity was not adequate, the electrode was pressed down into the hair and wetted with additional sterile saline solution.

C. Procedure

The P3 speller task involves a 6x6 grid of letters and numbers, as shown in the upper portion of figure 2, which are visually intensified at a rate of 125 msecs, flashing alternate rows and columns in a random presentation. The participant is instructed to count flashes of the letter or number that they are trying to spell[16].

In the alternate card game task, shown at the bottom of figure 2, the same attentional mechanism was used to select matching pairs of cards – the participants were first shown all of the cards and then they were flipped over and revealed one by one as the participant concentrated on them to form pairs. If a pair was not matched then both cards were flipped back over and the process started again.

The monitor was at a viewing distance of 80cm, causing it to subtend a visual angle of approximately 34 degrees. The participants did not use a mouse or keyboard and all interaction with the system was through the EEG.

The study was broken into two sessions. On the first day, subjects arrived and had the P3 speller task explained to them,

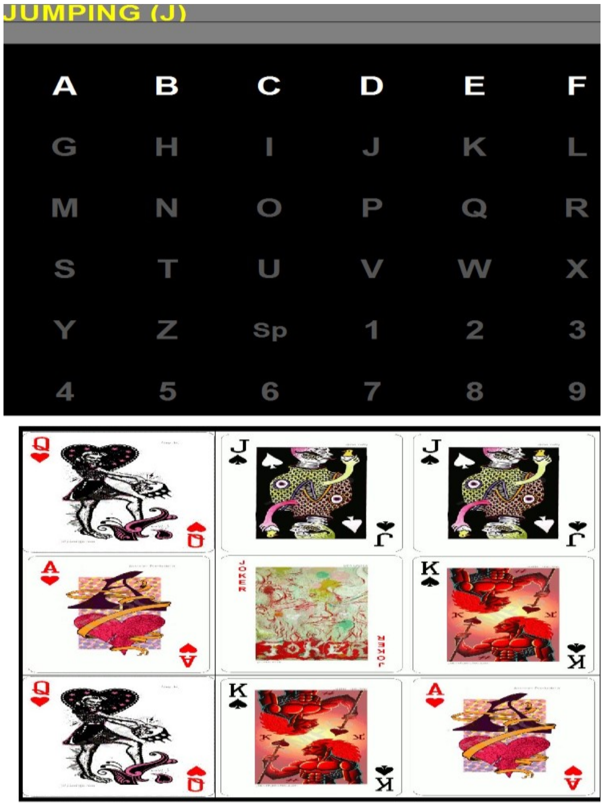

Figure 2: Interfaces for P3 speller task (top) and card game task (bottom)

then had the headset fit. Participants were told that each letter would be intensified for a moment, in rows and columns, and that they were to mentally count the number of times the letter they wished to spell turned bright. When participants understood, the first session was begun. This was with a very short phrase ("BIRD") in order to get the participant used to the task. Following this, they spelled 3 phrases ranging from 10 to 15 characters per phrase (e.g., "CHERRY BLOSSOMS").

Model training was attempted after the first and third phrase in order to maximize subject accuracy on the task. This training consists of using the prompted phrase paired with the results generated from the user. This alignment of known correct responses and user generated responses allows the system to train a linear classifier so as to be able to associate the user EEG inputs to more accurately predict user intention in the future trials. This training adjusts the inputs by electrode location and P300 timing to correctly weight the inputs on a per-user (and per-session) basis.

A fourth phrase was then given to test the fully trained model. The subject was then allowed to spell a word of their choosing (free spelling), writing it on paper before the experimenter began the session. A short break was then given (minimum two minutes, up to 5 minutes.) and then another free spelling trial. Finally, another prompted trial was given, and then the model was reverted to an untrained model and the initial phrase attempted again. All participants were given the same phrases, and the phrases were the same on both days (with the exception of the free spelling trials.)

Participants returned for a second session 24 hours later (plus or minus 3 hours). Procedure on day two was identical, with the exception that task instructions and the orienting phrase were not given. A questionnaire was filled out on both days asking for the participant's subjective effort ratings, comfort, and general alertness level.

A deviation from this procedure was that the 4 subjects who were run on the alternate card game task were tested for only one session. The card game results were not included in the statistical analysis for the training effect, but the card game represents a novel related paradigm which will be briefly discussed below.

IV. RESULTS

A. Data Analysis

The display of stimuli and real-time model is provided by the BCI2000 EEG framework[17]. This is a C++ framework that offers good performance and relatively easy customization. The package is only able to accommodate some fairly basic analyses within its own tools, though, so most of the detailed analysis was done within Matlab, using the EEGlab analysis package[18]. A few significant changes needed to be made to the data before they could be imported into EEGlab. The first was to map the event format of the results from the BCI2000 standard into a format amenable to analysis in EEGlab. A more significant theoretical point is that the epochs in the P3 speller in BCI2000 are overlapped, meaning that with an epoch length of 800 msecs and a display rate of 125 msec, each epoch contains multiple stimuli. This is not a problem, but does mean that the independent component analysis (ICA) features of EEGlab are not appropriate to use on this data set (since each epoch contains data from several stimuli and therefore have strong cross-dependencies.)[19] For this reason, all analyses below are in the time domain rather than in terms of more complex decompositions. All non-target epochs are those in which there is no target event anywhere in the epoch. The reference signal on the Emotiv also requires brief discussion - because the reference channel in the reversed electrode position is itself neurally active at location C4, the channels are all relative references to this position. In the case of a broadly distributed signal like the P300 this is not as convenient as an absolute signal but still allows reasonable data interpretation.

Three analyses were of primary interest. The first is of the P3 speller data on its own. Can we find a difference between target and non-target items that is significant, broadly distributed across electrode locations, and centered in the time range that we would expect a P300 response? Second, is there an effect of either fatigue or training (moving in opposite directions, of course)? And finally, is the alternate card game task generating similar patterns of data?

The question of training and fatigue requires teasing the effects apart carefully. One concern is that the task demand characteristics may not lead subjects to pay the utmost attention at all times. Because the classifier provides correct responses even when attention wavers, subjects noted verbally that it seemed easier on the second day. It would be preferable to have

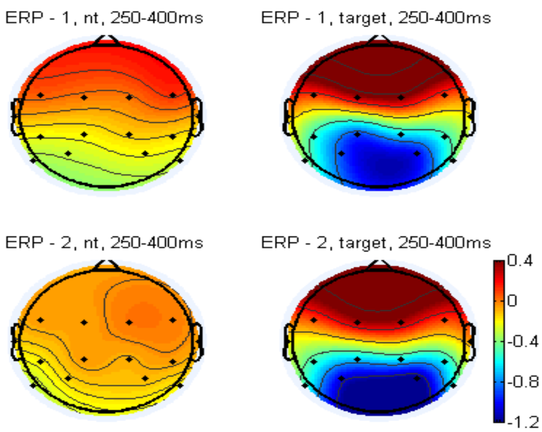

Figure 3: Topographical view of all electrode locations on day 1 (top) and day 2 (bottom) showing minimal activity on non-target trials (left) and a band of activity across the top of the head on target trials, between 250 and 400 msec in each epoch. Differences between target and non-target trials are significant at the p<.05 level.

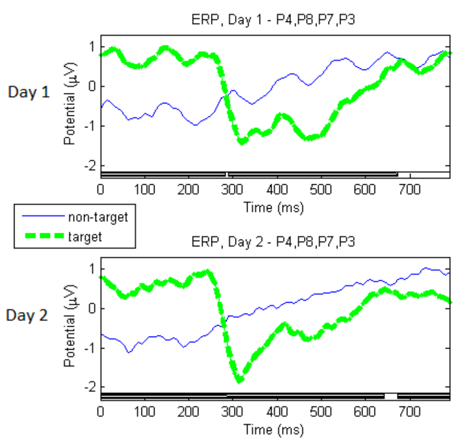

Figure 4: Non-target (blue) vs. target (green) ERP responses on day 1 (top) and day 2 (bottom). Dark bars at base of charts indicate regions where target and non-target are significantly different at the .05 level, showing a robust P300 signal on both days.

perceived difficulty remain constant and see a performance increase. As a partial way of negating this effect, we specifically looked at the 4 trials (two on day one, two on day two) where an untrained model was used on the same phrase; in these, the classifier does not confound motivational factors. Because one trial is at the beginning and one at the end of each day this also allows us a window onto attentional fatigue.

B. *Experimental Results*

The overall P300 effect is very robust in this experiment, across many trials of data (72 individual phrases in total). Figure 3 demonstrates the overall electrode positions and figure 4 the overall P300 at an average of electrode locations p3, p4, p7, and p8. Due to the central location (c4) of the Emotiv reference channel, the P300 was not accessible from the typical front-

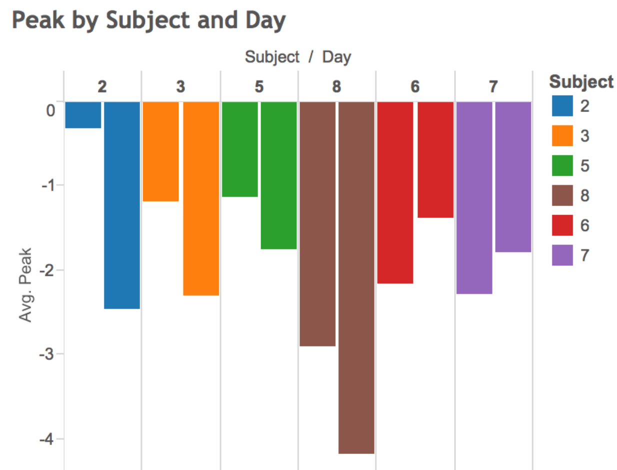

Figure 5: Average peak height on untrained model trials for day 1 (left bar in each pair) and day 2 (right bar in each pair) across subjects (coloured), showing the strength of the P300 response increasing through the training. Y-axis is in microvolts at the peak of the P300 response (range 250-400 msecs).

central channels. The posterior channels chosen here reflect the negative dipole of the positive frontal P300 component.

For the analysis of the training effect, the maximum peaks were picked in the interval of 250 msecs to 400 msecs post-stimulus-presentation, again on an average of the electrode locations at p3, p4, p7, and p8. These peaks differed significantly on day 1 vs day 2 with a day 1 mean of -1.66 microvolts and a day 2 peak of 2.31 microvolts (t-test on peak values shows p<.05). These results are broken out by subject in figure 5, showing a suggestive trend of improvement over time for most subjects. Subjects 6 and 7 began day 1 at a fairly high level of performance and this reversed on day 2, presumably due to motivational factors as they became bored with the task.

Overall accuracy rates ranged from 4% on the first untrained phrase to 90% on a phrase of their own choosing. On day 2, the first untrained phrase had an average of 10% correct. Because the number of subjects in the study was chosen to show

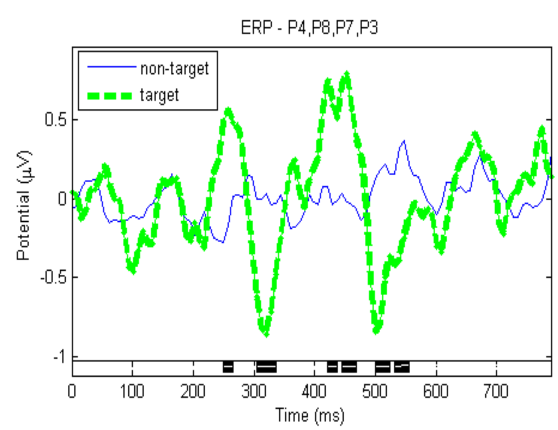

Figure 6: Target vs. non-target ERP responses at locations P3, P4, P7, P8 (averaged) on the card game task. The P300 for target presentations is significantly different than the non-target presentations at p<.05.

statistically significant results on individual P300 responses, the overall word error rate does not have enough samples to be statistically significant.

On the post-test questionnaire, subjects reported on headset comfort (1=very uncomfortable, 7=very comfortable) and saline solution irritation (1=very irritating, 7=no irritation) with average values of 6.2 and 7, respectively, indicating that headset comfort was not seen as much of an issue in this study, unlike in prior studies from other labs[13].

The card game task was treated in the same manner as the P3 speller data and shows a strong P300 peak (see figure 6). This shows that a similar visual attention mechanism is being used in the case of this game, although the cognitive parameters of the task differ.

V. Discussion

The two primary research questions in this paper are whether the Emotiv Epoc is sufficiently sensitive to detect ERPs, and whether we can use a P3 speller paradigm to show evidence of a P300 training effect (specifically strength increases). Our results show that the Emotiv, despite some limitations in the reference electrode location, does provide usable ERP data. Furthermore, the training effect over two sessions of P3 speller usage shows an overall clear and significant trend of better performance, though this trend did not exist for all individual participants.

It is likely that many of the factors leading to the successful use of P3 spellers are not strictly training of the brainwaves being produced, but rather training in how to put on the headset properly, how (and when) it is necessary to maintain attention, how to avoid movement artifacts, etc. Since the experimenter took care of headset fitting and gave instructions and feedback to minimize artifacts, this current study was not designed to show training in anything but the actual P300 brainwave production.

Anecdotally, one of the authors of this paper has been using the Emotiv to test various paradigm changes for many months and has seen large performance improvements during that time. This was one of the reasons to begin looking for a training effect. Of course, two sessions is not a lot of training time, but it is encouraging to see performance increases even in that short timeframe.

Past research has shown an effect of motivation on P300 responses[9]. This is supported by our participants' self-reported strategies. Most stuck with the "count the number of flashes" strategy described in the instructions, but other participants found ways of making the task more interesting. Two played word association games and one even inventing swear words that started with the letter they were looking for in an attempt to make it a stronger association. However, insufficient data were collected on alternate strategies to assess a performance impact.

In the card game, task difficulty can be manipulated by means of the card grid size, the number of pairs, and the time given to memorize them. The hope is that the cognitive challenge of the task can be manipulated while preserving the visual attention focus of the task. In longer term training studies, this independent manipulation of cognitive load versus attentional load seems very beneficial in keeping users interested. The discussion of P300 for gaming interfaces is directly relevant to this approach[8]. One problem with the current card game paradigm is that it does not function as a mechanism to do supervised model training – in the P300 speller, since the subject is being given the letters to spell, the system can align participant responses with the known correct responses in order to train the linear classifier to recognize responses better. In the case of the card game, the system does not know a priori which card is the correct response, so this training cannot occur. In this experiment, the subject was asked to verbally state which card they were trying to turn over so that the experimenter could keep track of "correct" responses for offline analysis, but other paradigms are being considered to alleviate this difficulty. Nonetheless, the fact that the P3 speller trials could be used to train a model for the card game task (allowing it to work at an effective degree of accuracy) is confirmatory evidence that the card game and P3 speller share a common mechanism.

Although the problem of the reference electrode being in a very active location and therefore causing the P300 to appear in negative fashion on uninvolved locations is cumbersome from a research perspective, it does still allow for reasonable signals to be captured. The current study does alleviate some concerns over the comfort and usability aspects of the Emotiv headset. Given the decreasing price of consumer EEG headsets and the strengthening evidence of a training effect, further study of these low-cost devices seems warranted and the authors are hopeful for wider adoption opportunities for EEG-mediated biofeedback approaches.

VI. Future Work and Conclusions

The Emotiv Epoc is a suitable device for ERP neurofeedback in settings where a full research EEG is cost-prohibitive. We see in this study some initial results indicating that brain training is effective in increasing the strength of P300 responses. We also show that an alternate paradigm, the card game, shows similar P300 effects; this or similar tasks may be more interesting than the P3 speller task for long-term ERP neurofeedback training.

Despite its suitability for training, the Emotiv electrode montage is not optimal for P300 studies: the reference channel is inconvenient and the saline-coupled electrodes are cumbersome. There is a new option forthcoming – the Emotiv Insight is a new capacitively coupled headset that will be evaluated. There are also some open source alternatives such as the OpenBCI project, which appears to result in good data, although the drivers to use it in BCI2000 are not yet available. Some other consumer headsets are recently on the market, such as the Interaxon Muse, but these do not seem to be easily modifiable to give access to the electrode locations that would be most useful for ERP studies. The predominant use of these consumer devices remains for frontal lobe slow wave activity, which because it is averaged over a much longer time is not as technically challenging to detect as ERPs.

With capable devices on the immediate horizon, though, it makes sense to work on strengthening the case for neurofeedback of ERP components as being effective and useful. Effectiveness seems straightforward to undertake through longitudinal studies and longer term use, but will require tasks which are intrinsically more interesting. Utility will need to tie those longitudinal uses to measures that correlate with known conditions. ADHD is the obvious first candidate, and the reason for the focus in this paper on visual attention, but as inexpensive and accessible headsets improve in capability, other ERP targets are also very interesting to consider.

VII. References

[1] T. W. Picton, “The P300 wave of the human event-related potential.,” Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, vol. 9, no. 4, p. 456, 1992.

[2] F. Aloise, P. Arico, F. Schettini, E. Lucano, S. Salinari, F. Babiloni, D. Mattia, and F. Cincotti, “Can the P300-based BCItraining affect the ERPs?,” International Journal of Bioelectromagnetism, vol. 13, no. 3, pp. 148–149, 2011.

[3] M. Marchetti and K. Priftis, “Effectiveness of the P3-speller in brain–computer interfaces for amyotrophic lateral sclerosis patients: a systematic review and meta-analysis,” Front Neuroeng, vol. 7, May 2014.

[4] D. Brandeis, T. H. van Leeuwen, J. Steger, K. Imhof, and H.-C. Steinhausen, “Mapping brain functions of ADHD children,” International Congress Series, vol. 1232, pp. 649–654, Apr. 2002.

[5] K. A. Kerns, K. Eso, and J. Thomson, “Investigation of a Direct Intervention for Improving Attention in Young Children With ADHD,” Developmental Neuropsychology, vol. 16, no. 2, pp. 273–295, Oct. 1999.

[6] S. Sutton, M. Braren, J. Zubin, and E. R. John, “Evoked-Potential Correlates of Stimulus Uncertainty,” Science, vol. 150, no. 3700, pp. 1187–1188, Nov. 1965.

[7] L. A. Farwell and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” Electroencephalography and Clinical Neurophysiology, vol. 70, no. 6, pp. 510–523, Dec. 1988.

[8] A. Kaplan, S. Shishkin, I. Ganin, I. Basyul, and A. Zhigalov, “Adapting the P300-Based Brain #x2013;Computer Interface for Gaming: A Review,” IEEE Transactions on Computational Intelligence and AI in Games, vol. 5, no. 2, pp. 141–149, Jun. 2013.

[9] H. M. Gray, N. Ambady, W. T. Lowenthal, and P. Deldin, “P300 as an index of attention to self-relevant stimuli,” Journal of Experimental Social Psychology, vol. 40, no. 2, pp. 216–224, Mar. 2004.

[10] S. K. Loo and S. Makeig, “Clinical Utility of EEG in Attention-Deficit/Hyperactivity Disorder: A Research Update,” Neurotherapeutics, vol. 9, no. 3, pp. 569–587, Jul. 2012.

[11] V. Prox, D. E. Dietrich, Y. Zhang, H. M. Emrich, and M. D. Ohlmeier, “Attentional processing in adults with ADHD as reflected by event-related potentials,” Neuroscience Letters, vol. 419, no. 3, pp. 236–241, Jun. 2007.

[12] M. Grierson and C. Kiefer, “Better brain interfacing for the masses: progress in event-related potential detection using commercial brain computer interfaces,” in Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, 2011, pp. 1681–1686.

[13] L. Mayaud, M. Congedo, A. Van Laghenhove, D. Orlikowski, M. Figère, E. Azabou, and F. Cheliout-Heraut, “A comparison of recording modalities of P300 event-related potentials (ERP) for brain-computer interface (BCI) paradigm,” Neurophysiologie Clinique/Clinical Neurophysiology, vol. 43, no. 4, pp. 217–227, Oct. 2013.

[14] M. Duvinage, T. Castermans, T. Dutoit, M. Petieau, T. Hoellinger, C. D. Saedeleer, K. Seetharaman, and G. Cheron, “A P300-based Quantitative Comparison between the Emotiv Epoc Headset and a Medical EEG Device,” 2012.

[15] https://emotiv.com/forum/forum4/topic1369/messages/ accessed April 13, 2015.

[16] http://www.bci2000.org/wiki/index.php/User_Reference:P3SpellerTask accessed April 13, 2015.

[17] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “BCI2000: a general-purpose brain-computer interface (BCI) system,” IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, pp. 1034–1043, June.

[18] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[19] M. X. Cohen, Analyzing Neural Time Series Data: Theory and Practice, 1 edition. Cambridge, Massachusetts: The MIT Press, 2014.