34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Poor performance in SSVEP BCIs: Are worse subjects just slower?

Christoph Guger, Brendan Allison, Christoph Hintermueller, Robert Prueckl, Bernhard Großwindhager, Christoph Kapeller, and Guenter Edlinger, *Member, IEEE*

Abstract— **Brain-computer interface (BCI) systems translate brain activity into messages or commands. BCI studies that record from a dozen or more subjects typically report substantial variations in performance, as measured by accuracy. Usually, some subjects attain excellent (even perfect) accuracy, while at least one subject performs so poorly that effective communication would not be possible with that BCI. This study aims to further explore the differences between the best and worst performers by studying the changes in estimated accuracy within each trial in an offline simulation of an SSVEP BCI. Results showed that the worst performers not only attained lower accuracies, but needed more time after cue onset before their accuracies improved substantially. This outcome suggests that poor performance may be partly (though not completely) explained by the latency between cue onset and improved accuracy.**

I. Introduction

A brain-computer interface (BCI) system is a device that translates a user's brain activity into messages or commands in a closed-loop system with feedback. These messages or commands might be used to control a variety of devices, such as spelling systems, smart home devices, or prosthetic limbs [1, 10, 17, 18]. This article focuses on non-invasive BCIs, which comprise the substantial majority of BCI publications [12].

Among non-invasive BCIs, three types of brain signals are typically used for control. Some BCIs rely on Event-Related desynchronization (ERD) activity associated with imagined movements [8], others rely on P300 signals elicited whenever a target item flashes [9], and a third category relies on SSVEP activity resulting from attention to oscillating objects [14].

With any BCI, improving accuracy is a major goal. For reasons that are not fully understood, some people have trouble using BCIs well, and a small minority of subjects cannot use BCIs at all [23]. In previous work, we collected data from a large number of BCI users chosen from the general public, hoping to explore how many people could attain different accuracies. With ERD and P300 research, these efforts successfully identified performers at different skill levels [8, 20]. However, the only large-scale effort with SSVEP used substantially different methods, and did not record EEG, leaving no data available for further analysis [22]. Moreover, while all of these studies identified people who did not perform well, none of them explored the resulting EEG data to identify why some subjects performed poorly.

One way to explore causes of poor SSVEP BCI performance is to analyze the changes in their performance over time. Typically, articles that estimate or analyze BCI performance within a trial only report the peak accuracy within that trial. Although this approach is common in the literature [4, 13], it may obscure important dynamics within each trial [3, 6]. Other factors such as the mean accuracy, the latency between cue onset and acceptably accurate classification, and the accuracy at the end of the trial might reveal information about brain dynamics across different users that cannot be detected via peak accuracy measurement alone.

Hence, we have been exploring the estimated accuracy within a trial from a large number of subjects with an online simulation of an SSVEP BCI. To date, 53 subjects participated in the paradigm described below, and more data may be collected. Our goal here is not to report on the complete results, which would be beyond the scope of this paper. Instead, we wish to compare the time courses of the error rates within each trial between the best and the worst performers, to determine whether paradigmatic changes might improve performance among poor performers.

II. Methods

53 healthy people volunteered for this study (35 male). The age range was 8-73, mean age 28.72 ± 12.6. All subjects had normal or corrected to normal vision, and provided informed consent before beginning the study. There was one subject under 18, who participated with her parents' consent. None of the subjects had prior experience with a BCI. Thus, the subjects were a fairly reasonable subset of the healthy population.

At the beginning of the recording session, each person was prepared for recording using g.BUTTERFLY active electrodes. These electrodes require a small amount of electrode gel, and do not require skin preparation. Fig. 1 shows the electrode montage used in this study. Data were recorded from eight posterior electrode sites positioned according to the International 10-20 system, with a reference electrode on the right earlobe and a ground electrode over site FPz. Data were sent to a g.USBamp amplifier sampling at 256 Hz, with a bandpass filter of 0.5-30 Hz and a notch filter at 50 Hz.

This research was supported by the EU Projects BrainAble (Project number 227447), VERE (Project number 257695), Decoder (Project number 247919), as well as the multinationally funded ALIAS project.

C. Guger, C. Hintermueller, R. Prueckl, B. Großwindhager, C. Kapeller, and G. Edlinger are with g.tec medical engineering GmbH, Guger Technologies OG, Herbersteinstrasse 60, 8020 Graz, Austria

B. Z. Allison is with the Department of Cognitive Science, University of California at San Diego, La Jolla, California, USA.

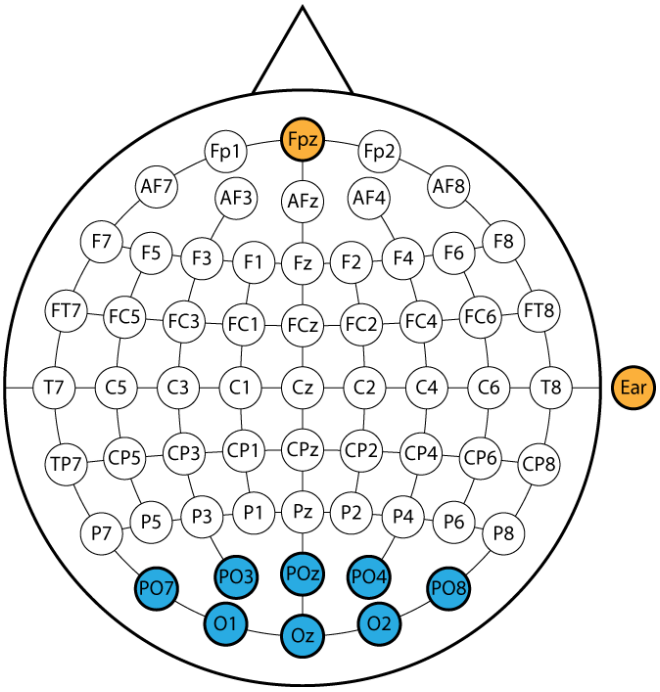

Figure 1. The electrode montage used in this study. The eight recording sites are located over posterior regions of the head, with a ground on FPz and reference electrode on the right earlobe.

After being prepared for EEG recording, subjects participated in a training run to collect the data needed to train the classifier, then a test run. Subjects viewed a g.SSVEPbox with four LEDs, on the top, right, bottom, and left (See Fig. 2). During each run, subjects first had five trials in which the top LED was the target, followed by five trials each with right, bottom, and left targets. The order of these twenty trials was the same for all subjects. Each run began with a ten second delay before the first trial. Each trial began with a three second pause. Next, a small green light appeared on the box about 2 mm from one of the LEDs. Subjects were asked to focus their attention to the LED closest to that green light. Simultaneously, the four LED began to oscillate at 10 Hz (top box), 11 Hz (right box), 12 Hz (bottom box), or 13 Hz (left box).

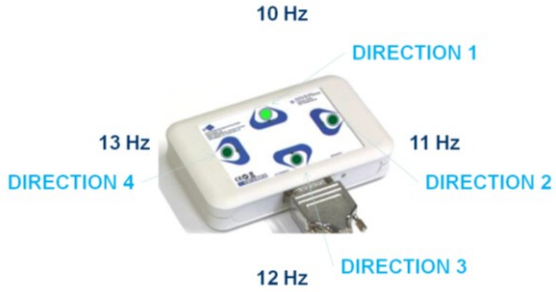

Figure 2. The g.SSVEP box used to present stimuli. Each of the four LEDs flickered at a different frequency from 10-13 Hz.

Subjects were asked to focus on the target LED for seven seconds, after which the trial ended and the lights on the g.SSVEPbox turned off. Next, the subjects had a short break between runs, during which the classifier was trained on their data. Then, subjects participated in a second run. This run was identical to the training run, except that the BCI estimated the accuracy online during each trial. Feedback was given in form of a number indicating the selected item. A minority of subjects chose to volunteer for one or two additional online runs.

All data presentation, recording, and analysis was managed by g.BCIsys as shown in Figure 3. g.BCIsys uses Simulink as a rapid prototyping platform to run real-time experiments [7]. The top half of Fig. 3 summarizes the data collection and analysis approach. The bottom of Fig. 3 shows how a g.STIMbox was connected to the g.SSVEPbox to help generate the stimuli.

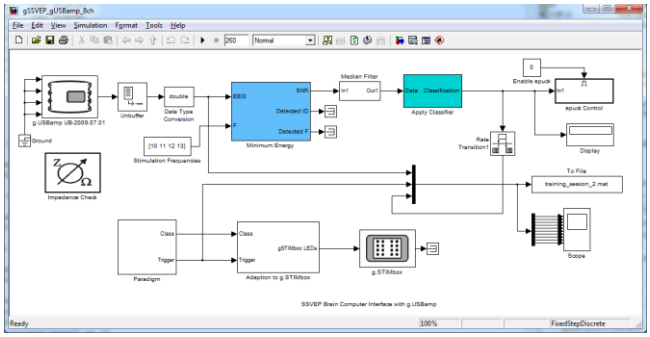

Figure 3. The Simulink model used for the SSVEP study.

The pattern recognition procedure began with the minimum energy combination approach to determine the spatial filter settings that would yield the best signal-to-noise ratio or SNR [21]. Next, the system estimated the SNR using a Levinson AR Model (order 7) based on the previous 768 sample points (3 seconds). The software updated this classification every 200 ms. A linear discriminant analysis (LDA) classifier was used for real-time pattern classification based on the SNR [7].

III. Results

A. General results

The mean peak accuracy (the average of each subject's best performance) is 94%. Chance accuracy in this four-class paradigm would be 25%. 31 of the 53 subjects attained 100% accuracy. The four best subjects were defined as the four subjects who attained 100% accuracy in the shortest possible time after cue onset. Four subjects attained 100% accuracy within four seconds of cue onset, and retained 100% accuracy throughout the remainder of the trial.

The worst four subjects attained between 70-80% peak accuracy. Two of them attained about 70% accuracy, and the next two attained 75-80% accuracy.

B. Four best subjects

Fig. 4 shows the results for the four best subjects in this study. Like all subjects' results, error is high during the rest period before cue onset. In these four subjects, error declines substantially within one to two seconds after cue onset.

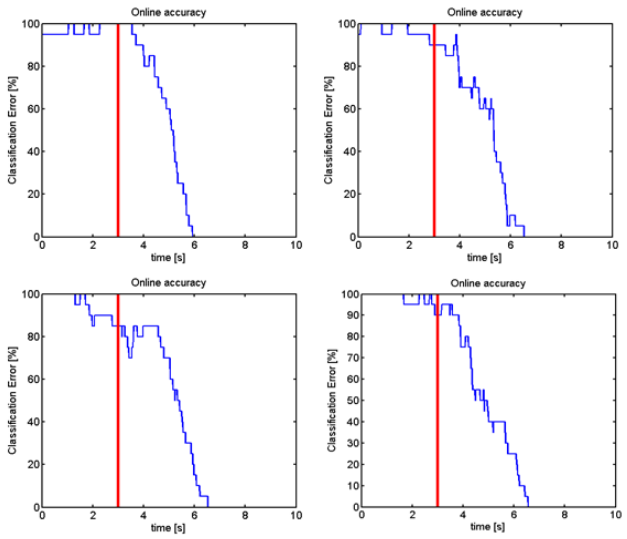

Figure 4. The four subjects who performed best. All subjects' estimated error dropped to zero within four seconds of cue onset, and remained zero throughout the rest of the trial. The vertical axis shows classification accuracy, and the horizontal axis shows the time from trial onset. The vertical red line at 3 seconds indicates the onset of the cue, which is also when the four different LEDs began oscillating.

C. *Four worst subjects*

Fig. 5 shows the results for the four worst subjects in this study. Although these subjects still show a reduction in error rates after cue onset, it tends to occur more than two seconds after cue onset. Furthermore, three of these four subjects' error rates continue to decline at the end of the trial, implying that peak accuracy might improve with longer trials.

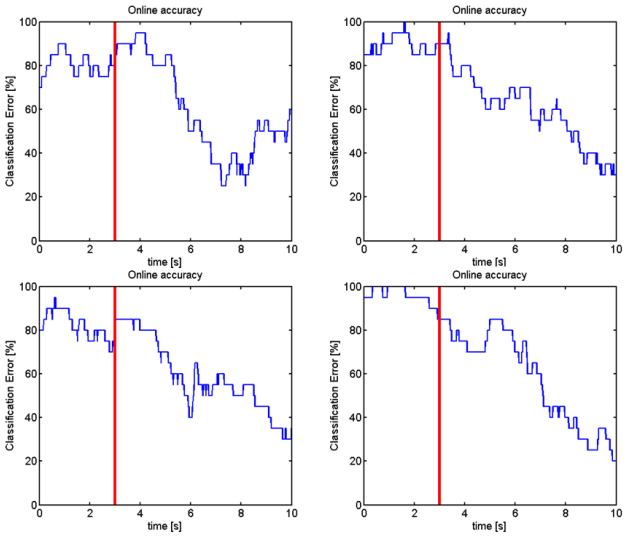

Figure 5. The four subjects who performed worst. Their peak accuracy was between 70-80%. The axes and the red line are described in the preceding figure caption.

IV. Discussion

The peak accuracy may be less important in realworld studies. Since a BCI does not know which time segment contains the peak accuracy, the likelihood of error depends much more on the mean than peak accuracy. In addition, reporting the mean accuracy in synchronous BCIs such as these could help readers estimate control in a continuous task such as virtual movement, when maintaining high accuracy over seven seconds or longer may be critical [2,3].

A. *Latency: A major delay*

Speed is a critical factor in any communication system. In some BCIs, one critical factor that limits speed is the latency between the onset of a cue that directs or helps subjects to perform a specific mental task and the moment that the resulting brain activity can be effectively classified. This latency may be impossible to measure, or meaningless, in many situations, such as with asynchronous BCIs [3, 4, 14, 17, 18, 19] or whenever users freely choose when to change mental tasks, since inferring the moment of decisionmaking can be difficult [11].

This latency reflects several stages. The cue (in this case, a green light) must travel through the visual system. The subject must decide how to respond, then shift attention to the target stimulus and engage attention there [16]. Thalamocortical oscillations must lead to sufficient synchronous EEG activity that the classifier used in this study could attain an acceptable accuracy level [5, 6, 15].

In this BCI, the latency was over two seconds for the best subject (shown in the top left panel of figure 3), assuming a reasonable accuracy threshold of 70-80% [1, 3, 12, 17, 18]. This is a substantial delay in any communication system, and further research should explore how to reduce latency in different BCIs. Training may help, at least with the later stages, as well as alternate interfaces that allow subjects to anticipate their next mental activity. For example, a subject who uses SSVEP or other activity to navigate a real or virtual environment could often anticipate the next intended movement command [2, 19]. In the present study, subjects very probably learned to anticipate the next target, but the LEDs were not active until cue onset. Alternate sensors, stimuli, classification software, interfaces, and other factors may also help.

B. *Best vs. worst performers*

As with all BCI studies, some subjects in the present study performed much worse than other subjects, with performance defined by low peak accuracy based on offline analysis. We found that our worst subjects differed from the best subjects in two other ways. First, the worst subjects typically required a longer latency before their error declined significantly. Second, the worst subjects continued to improve throughout the trial, and therefore a longer trial might be advantageous.

Of course, subjects who attain 100% performance cannot improve further. However, this result could be relevant for poor performers, because it implies that longer trials might yield better performance.

An alternate outcome was possible, since our analyses might have revealed that poor performers did not improve throughout the trial. In this case, longer trials would not have helped. Instead, the present study suggests greater attention to the dynamics of brain activity, and associated accuracy

measures, within (at least) SSVEP BCI systems. Furthermore, the results suggest that, to some extent, worse subjects might simply be slower. That is, the difference in accuracy between good and bad performers might be reduced with longer trials. It is also conceivable that training poor performers could help both accuracy and latency. If performance improves with longer trials, then poor performers might learn how to focus attention better within a shorter time. However, latency only partly explains poor performance. Worse subjects are not just slower.

This work was based on offline analyses. Although these analyses can provide reasonable estimates of online performance, using methods similar to those described here [3], we recommend following up with an online BCI. Another critical future direction involves patients. The present study only used healthy subjects, while patients may have different performance dynamics due to visual deficits, medication, fatigue, neuropsychiatric conditions such as hemineglect, etc.

V. Conclusion

This work leads to three general conclusions, which are significant for different reasons. The first reason is methodological: reporting only peak accuracy within a trial may obscure important changes within a trial. This could encourage more thorough reporting in the future, and foster more in-depth analysis of BCI performance. Second, poor performers' error rates take longer to decline after cue onset, perhaps in part because they require more "charging time" to develop the thalamocortical oscillations needed for accurate SSVEP BCI use. This could encourage future research in to why some people perform poorly with SSVEP BCIs. Third, poor performers' error rates often continue to decline at the end of the trial. Hence, longer trials could help such users.

References

[1] Allison, B. Z., Wolpaw, E. W., and Wolpaw, J. R. (2007). Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, *4*(4), 1-12.

[2] Allison, B. Z., Faller, J., and Neuper, C. (2012). BCIs that use steady-state visual evoked potentials or slow cortical potentials. J. Wolpaw & E. W. Wolpaw (Eds.), *Brain-Computer Interfaces: Principles and Practice*, Oxford Univ. Press, Oxford: 241-250.

[3] Billinger, M., Daly, I., Kaiser, V., Jin, J., Allison, B. Z., Müller-Putz, G. R., and Brunner, C. (In press.) Is it Significant? Guidelines for Reporting BCI Performance. In Allison, B. Z., Dunne, S., Leeb, R., Millán, J., and Nijholt, A. (Eds.), *Toward Practical BCIs,* Springer Verlag, London.

[4] Brunner, C., Allison, B. Z., Altstätter, C., and Neuper, C. (2011). A comparison of three brain-computer interfaces based on event-related desynchronization, steady state visual evoked potentials, or a hybrid approach using both signals. *Journal of neural engineering*, *8*(2), 025010. doi:10.1088/1741-2560/8/2/025010

[5] Burkitt, G. R., Silberstein, R. B., Cadusch, P. J., and Wood, A. W. (2000). Steady-state visual evoked potentials and travelling waves. *Clinical Neurophysiology* 111, 246-258.

[6] Daly, I., Grissmann, S., Brunner, C., Allison, B. Z., and Müller-Putz, G. R. Hybrid BCI classification via dynamic re-weighting *Presented at the TOBI workshop 2012*

[7] Guger, C., Schlögl, A., Neuper, C., Walterspacher, D., Strein, T., and Pfurtscheller, G. (2001) Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9, 49-58.

[8] Guger, C., Edlinger, G., Harkam,W., Niedermayer, I., and Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain-computer interface (BCI)? *IEEE Transactions on Neural Systems and Rehabilitation Engineering 11*, 145-147. doi:10.1109/TNSRE.2003.814481

[9] Jin, J., Allison, B. Z., Wang, X., and Neuper, C. (2012). A combined brain-computer interface based on P300 potentials and motion-onset visual evoked potentials. *J. Neurosci. Methods* 205(2), 265-276.

[10] Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., and Birbaumer, N. (2001) Brain-computer communication: unlocking the locked in. *Psychol. Bull.* 127(3), 358-375.

[11] Libet, B., Wright, E. W. Jr., Feinstein, B., and Pearl, D. K. (1979) Subjective referral of the timing for a conscious sensory experience: a functional role for the somatosensory specific projection system in man. *Brain* 102(1), 193-224.

[12] Mason, S. G., Bashashati, A., Fatourechi, M., Navarro, K. F., and Birch, G. E. (2007) A Comprehensive Survey of Brain Technology Interface Designs. *Annals of Biomedical Engineering* 35(2), 137-169.

[13] Neuper, C., Scherer, R., Reiner, M., and Pfurtscheller, G. (2005) Imagery of motor actions: differential effects of kinesthetic versus visual-motor mode of imagery on single-trial EEG. *Brain Research Cognitive Brain Research 25*, 668-677.

[14] Ortner, R., Allison, B. Z., Korisek, G., Gaggl, H., and Pfurtscheller, G. (2011). An SSVEP BCI to control a hand orthosis for persons with tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19, 1-5.

[15] Pfurtscheller, G., and da Silva, F. H. L. (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology* 110, 1842-1857.

[16] Posner, M. I., and Petersen, S. E. (1990) The attention system of the human brain. *Annual Review of Neuroscience* 13, 25-42.

[17] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G.,and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology* 113, 767-791.

[18] Wolpaw, J. R., and Wolpaw, E. W. (2012). Brain computer interfaces: Something new under the sun. In J. Wolpaw & E. W. Wolpaw (Eds.), *Brain-Computer Interfaces: Principles and Practice* (pp. 3-14). Oxford: Oxford Univ. Press.

[19] Millan J del R., Renkens, F., Mourino, J. and Gerstner, W. (2004)- Noninvasive brain-actuated control of a mobile robot by human EEG *IEEE Trans. Biomed. Eng*. 51, 1026–33.

[20] Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., Gramatica, F. and Edlinger, G. (2009). How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters* 462(1), 94-98.

[21] Friman, O., Volosyak, I. and Gräser, A. (2007). Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE Trans. Biomed. Engineering* 54, 742-750.

[22] Allison, B.Z., Valbuena, D., Lueth, T., Teymourian, A., Volosyak, I., and Gräser, A. (2010). BCI demographics: How many (and what kinds of) people can use an SSVEP BCI? *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 18(2): 107-116.

[23] Allison, B.Z. and Neuper, C. (2010). Could anyone use a BCI? In: Applying our Minds to Human-Computer Interaction, Tan, D.S. and Nijholt, A. (Eds.) *Brain-Computer Interfaces. Human-Computer Interaction Series,* Springer Verlag, London, 35-54.