 DOI:10.3217/978-3-85125-260-6-185

Towards a Broad-Scale Usability Evaluation of Hybrid BCIs

R. Lorenz[1], J. Pascual[1], B. Blankertz[2], C. Vidaurre[1]

[1]*Machine-Learning Group;* [2]*Neurotechnology Group, Berlin Institute of Technology*

Correspondence: R. Lorenz, Berlin Institute of Technology, Berlin, Germany. E-mail: lorenz.romy@gmail.com

Abstract. For the study, three graphical user interfaces (GUIs) were designed for their prospective use in controlling a brain-computer interface (BCI)-driven upper-limb neuroprosthesis. The action selection was divided into two stages: selection and confirmation that were controlled using event-related potentials (ERP) or motor imagery (MI). By evaluating usability on a broad-scale including behavioural, subjective and EEG data, the study provides valuable insights into the underlying dynamics that cause the differences in performance across the GUIs.

Keywords: Usability, User-Centered, Ease of Use, Hybrid BCI, Motor Imagery, ERP, Neuroprosthetics

1. Introduction

In contrast with the tremendous increase of usability evaluations in the field of human-computer interaction, the awareness of the importance of a user-centered perspective in the field of BCI assistive devices for patients is only growing moderately. Most BCI systems are exclusively evaluated in terms of classification accuracy and speed [Pasqualotto et al., 2012]. Besides including these common efficiency measures, the evaluation of further usability aspects such as ease of use, learnability and workload could improve user efficiency and satisfaction [Plass-Oude Bos et al., 2011]. Nonetheless usability is typically evaluated by means of questionnaires and behavioural data, it is a "golden opportunity" to extract usability-related features from the brain by recording the same neurophysiological signal the BCI is controlled with [van de Laar et al., 2011]. Within the context of the MUNDUS project [Pedrocchi et al., 2010], three different BCI GUIs were proposed for a prospective use in controlling a neuroprosthesis. The present work strikingly demonstrates the benefit of a broad-scale methodology by evaluating the usability of the interfaces based on results obtained from various data sources.

2. Material and Methods

Twelve healthy subjects (6 female; mean age: 26.2 ± 2.9 years) took part in the study. Brain activity was recorded using 64 electrodes placed according to the international 10-20 system. Three different GUIs were presented to each subject. The order of the GUIs was counterbalanced across the participants. The task for each GUI consisted of a two-stage action selection. First, subjects selected one of six symbols representing possible actions executed by a neuroprosthesis, and then they had to confirm or cancel this selection. For the experiment, a solely ERP-based and two hybrid combinations were tested: (1) selection with ERP, confirmation with ERP (ERP-ERP), (2) selection with ERP, confirmation with MI (ERP-MI) and (3) selection with MI, confirmation with ERP (MI-ERP). The ERP paradigm for the selection and confirmation stage was derived from the Center Speller [Treder et al., 2011]. For more details about the GUI design see [Pascual et al., 2013]. For the assessment of usability aspects, the NASA-TLX (workload) and the *use quality* dimension (ease of use and learnability) of the User Experience Questionnaire (UEQ) were administered after each GUI.

3. Results

Table 1. *Behavioural and questionnaire results of each GUI and corresponding statistics.*

	ERP-ERP	*ERP-MI*	*MI-ERP*	*F-value*	*P-Value*
Selection Accuracy	98.46 %	96.55 %	83.47 %	$F_{(2, 18)} = 6.315$	$p = .025$*
Confirmation Accuracy	96.26 %	93.38 %	92.24 %	$F_{(2, 18)} = 1.061$	$p = .367$
NASA-TLX (scale: 0 to 100)	28.83	35.00	62.50	$F_{(2, 18)} = 19.627$	$p < .001$*
Use Quality (scale: -3 to +3)	1.83	1.33	0.79	$F_{(2, 18)} = 4.913$	$p = .020$*

Behavioural (accuracy) and questionnaire results (Table 1) were analyzed using one-way repeated measures ANOVAs (α-level: 0.05). For offline ERP analysis, the filtered and down-sampled signal was divided into overlapping epochs ranging from -200 to 800 ms relative to the onset of the stimulus. As a measure of discriminability of target vs. nontarget, the *sgn* r^2 was computed for all channels. The grand average is shown in Fig. 1 for the comparisons of the selection and confirmation stage.

4. Discussion

Results clearly indicate that the ERP-ERP GUI surpasses both hybrid approaches in terms of effectiveness. However, the lower accuracies cannot solely be traced back to the MI mode itself. For the confirmation stage, the ERP-MI GUI is even more accurate than the MI-ERP GUI (see Table 1). Although this finding is not of statistical significance, it is unexpected insofar as the exact same paradigm was applied for the ERP-ERP and MI-ERP GUI. Observations of a lower P3 discriminability for the hybrid GUIs (see Fig. 1) provide a neurophysiological explanation since the online classification depends on the discriminability of the P3. Elucidation to the neurocognitive dynamics accounting for the lower P3 discriminability and the ensuing lower performance could be found in our questionnaire data. Workload and use quality scores seem to correlate with the discriminability of the P3: the higher/lower the workload/use quality score, the lower the P3 discriminability. Studies [Cox-Fuenzalida et al., 2006] showing how sudden changes in workload drastically impair performance confirm our observations made for the switch from the mentally high loading MI selection paradigm to the lower loading ERP-based confirmation. Other studies [Reuderink et al., 2009] point towards frustration as having a detrimental impact on BCI performance (as reflected in the low use quality score for the MI-ERP GUI). Nevertheless, the utility of *sgn* r^2 as e.g. a workload or frustration index remains open and needs further investigation. In any case, the broad-scale methodology of the present study proved to provide valuable insights into the underlying dynamics causing the performance differences between the GUIs.

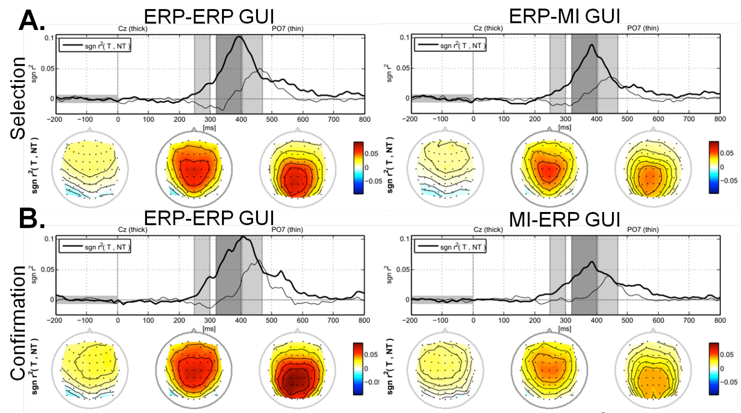

Figure 1. *Graphics A. and B. plot the time evolution of sgn* r^2 *between targets and nontargets for electrodes Cz and PO7 in the first row and the scalp plots for the grey shaded time intervals in the second. A. compares the selection stage between the ERP-ERP and ERP-MI GUI while B. compares the confirmation stage between the ERP-ERP and MI-ERP GUI.*

Acknowledgements

This work was supported by ICT-248326 and BMBF grant 01GQ0850. This publication only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

References

Cox-Fuenzalida L-E, Beeler C, Sohl L. Workload history effects: A comparison of sudden increases and decreases on performance. *Curr Psychol*, 25(1):8–14, 2006.

Pascual J, Lorenz R, Blankertz B, Vidaurre C. Hybrid EEG-based User Interface for Action Selection. In *Conf Proc ICNR*, 1171-1176, 2013.

Pasqualotto E, Federici S, Belardinelli M. Toward functioning and usable brain-computer interfaces (BCIs): A literature review. *Disabil Rehabil*, 7(2):89–103, 2012.

Pedrocchi A et al. MUNDUS: MUltimodal Neuroprosthesis for Daily Upper limb Support. Proc BioMed@POLIMI:1–4, 2010.

Plass-Oude Bos D, Gürkök H, van de Laar B, Nijboer F, Nijholt A. User Experience Evaluation in BCI: Mind the Gap! *Int J Bioelectromag*, 13(1):48–49, 2011.

Reuderink B, Nijholt A, Poel M. Affective Pacman: A Frustrating Game for Brain-Computer Interface Experiments. In Nijholt A, Reidsma D, Hondorp H (eds.). Intelligent technologies for interactive entertainment, 221–227, 2009.

Treder MS, Schmidt NM, Blankertz B. Gaze-independent BCIs based on covert attention and feature attention. *J Neural Eng*, 8(6):66003, 2011.

van de Laar B, Gürkök H, Plass-Oude Bos D, Nijboer F, Nijholt A. Perspectives on User Experience Evaluation of Brain-Computer Interfaces. In Stephanidis C (ed). *Universal Access in Human-Computer Interaction. Users Diversity*, 600–609, 2011.