"Toward a Usability Evaluation of BCIs"

Pasqualotto, Emanuele ; Simonetta, A ; Gnisci, Veronica ; Federici, S ; Olivetti Belardinelli, M

ABSTRACT

If BCIs could be considered and evaluated as assistive technology facilitating life activities, they could avoid dissatisfaction and prevent abandonment. We tested two keyboard controlled Java BCI prototypes, based on the Thought Translation Device and the P300 Speller, on 61 participants with different computer skills performing a Copy Spelling Task. We then administered usability and cognitive workload questionnaires. The results have shown significant differences in the number of performed errors, in satisfaction, and in the cognitive workload invested in the task. We found that the Thought Translation Device was more error-resistant, less stressful, and more satisfactory for the users.

CITE THIS VERSION

Toward a Usability Evaluation of BCIs
Emanuele Pasqualotto, Alessandro Simonetta, Veronica Gnisci, Stefano Federici, Marta Olivetti Belardinelli

Abstract. If BCIs could be considered and evaluated as assistive technology facilitating life activities, they could avoid dissatisfaction and prevent abandonment. We tested two keyboard controlled Java BCI prototypes, based on the Thought Translation Device and the P300 Speller, on 61 participants with different computer skills performing a Copy Spelling Task. We then administered usability and cognitive workload questionnaires. The results have shown significant differences in the number of performed errors, in satisfaction, and in the cognitive workload invested in the task. We found that the Thought Translation Device was more error-resistant, less stressful, and more satisfactory for the users.

Keywords: BCI, Usability, Cognitive Workload, NASA TLX, SUS

1. Introduction
Since BCIs could provide paralysed people a means of communication, this technology should be considered and treated as an assistive technology, which facilitates life activities [Pasqualotto et al., 2009]. Therefore, BCIs need an assessment process to match person and technology, in order to avoid dissatisfaction and abandonment [Scherer, 2000]. The tradition of Human Computer Interaction has already given us most of the tools needed to analyse and evaluate technology; although researches aimed at evaluating human factors in BCIs’ design are still scarce. In this study we aim at assessing the usability of two BCI prototypes by measuring the interaction with the systems in context, considering the performance, cognitive workload, and satisfaction of non-disabled users.

2. Material and Methods
2.1 Interface prototype
In order to evaluate only visual layout features, removing all the possible EEG-related issues, we developed two keyboard controlled Java prototypes. We decided to investigate the Language Support Program used in the Thought Translation Device (TTD) [Neumann, et al., 2003] and the P300 Speller (P3S) [Sellers and Donchin, 2006]. While the TTD is based on Slow Cortical Potentials and allows users to select between two alternatives (e.g. yes/no), the P3S is based on P300 event related potential and allows users to use a flashing 6 × 6 matrix of characters. We developed the prototypes in order to be controlled using a PC keyboard, considering the original nature of the input: TTD is controlled by using 2 keys, one for each half of the alphabet, P3S is controlled by using one key to be selected when the target appears. We chose these two interfaces because, due to the EEG components used, they represent two opposite ways of using an input in a speller device: dichotomic versus multi-choice.

2.2 Procedure
We tested our BCI prototypes on 61 participants (31 F, 30 M; age: mean 22.3, min 18, max 38) with different computer skills, randomly assigned to one of two groups, one for each prototype.
Using a Copy Spelling Task (CST), wherein users are asked to write on the screen through the interface, our participants wrote in one session a list of ten randomized words (111 characters) previously selected by means of a frequency list of words in Italian written language. We then administered the ‘System Usability Scale’ (SUS) [Brooks, 1996] in order to measure usability and the NASA TLX [Hart and Staveland, 1988] in order to measure the cognitive workload due to the rigid time constrain of the P3S flashing matrix compared to the constrain-free use of the TTD, bit rates were not comparable and we did not use it as a measure.

3. Results

We analysed the data using a GLM procedure and a $2 \times 2 \times 3$ design, where the variables were the Interface (PS3 and TTD), Gender, and Computer Skills (Low, Medium, High). We used the Number of Errors as a measure; moreover, we differentiated between two kinds of error: a choice-based error and a recognition-based error. While the first one represents a wrong choice, a choice that led to the wrong result, the second one is a lack of recognition of the letter to select, which led to pursue the research of the desired target. We used the SUS and NASA TLX values as measures. We also analysed each single component of the NASA TLX. Based on recent findings [Borsci et al., 2009] we analysed SUS as a global scale and as a two-factor scale, Usability and Learnability. Gender and Computer Skills did not interact with the measures, showing that performance and satisfaction are not influenced by users' personal characteristics. On the contrary, the Interface played a major role, interacting with the Number of Errors ($F(1,60) = 24.8, p < .00$; LSP: M 9.87, SD 8.76; P3S: M 32.35, SD 21.58), with the Usability score ($F(1,60) = 8.7 p = .005$; LSP: M 68.25, SD 15.77; P3S: M 55.00, SD 17.26), and with the cognitive workload invested in the use ($F(1,60) = 4.08 p = .049$; LSP: M 43.01, SD 20.93; P3S: M 56.11, SD 26.12). Moreover, we found significant differences in the score of the wrong choice ($F(1,60) = 30.1 p < .00$; LSP: M 5.70, SD 4.80; P3S: M 26.84, SD 19.23) and in the NASA-TLX Frustration score ($F(1,60) = 5.9 p = .019$; LSP: M 6.39, SD 8.95; P3S: M 13.83, SD 13.19). The analysis of the two factors of SUS confirmed a difference in the Usability score ($F(1,60) = 10.76, p = .002$; LSP: M 52.42, SD 13.32; P3S: M 40.48, SD 14.02) but not in the Learnability score.

4. Discussion

Although several years have passed since the beginning of BCI research, there is still a lack in the use of usability and psychological methods to evaluate this technology. Even though an improvement is required, our methodology allowed us to begin an exploration of the usability of BCIs. The Thought Translation Device represents a binary modality that we can find even in sensory-motor-rhythm based BCI, in contrast with the multi-choice possibility of the P300 Speller. Disregarding any possible EEG-based issue, the TTD allowed our participants to be more accurate, reducing the possibility of error. Recognition is usually considered easier than recollection, and our data are consistent with this assumption. In fact, we did not find any difference in the recognition-based errors, while we found them in the choice-based error. Consistently with the error scores, NASA-TLX has shown that participants found the TTD less stressful and less frustrating. In the same way, the SUS global scores and the Usability factor have shown that the TTD is generally considered preferable.

References


