Workload measurement in a communication application operated through a P300-based brain–computer interface

2011 J. Neural Eng. 8 025028

(http://iopscience.iop.org/1741-2552/8/2/025028)

J. Neural Eng. **8** (2011) 025028 (6pp) doi:10.1088/1741-2560/8/2/025028

Workload measurement in a communication application operated through a P300-based brain–computer interface

A Riccio[1,2,8], **F Leotta**[1,3], **L Bianchi**[1,4], **F Aloise**[1], **C Zickler**[5], **E-J Hoogerwerf**[6], **A Kübler**[5,7], **D Mattia**[1] **and F Cincotti**[1]

[1] Neuroelectrical Imaging and BCI Lab, Fondazione Santa Lucia IRCCS, Rome, Italy
[2] Department of Psychology, University of Rome 'La Sapienza', Rome, Italy
[3] Dipartimento di Informatica e Sistemistica 'A Ruberti', 'Sapienza' Università di Roma, Italy
[4] Dipartimento di Neuroscienze, Università Tor Vergata, Rome, Italy
[5] Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Tübingen, Germany
[6] Ausilioteca, AIAS Bologna onlus, Bologna, Italy
[7] Department of Psychology I, University of Würzburg, Würzburg, Germany

E-mail: a.riccio@hsantalucia.it

Received 1 December 2010
Accepted for publication 11 February 2011
Published 24 March 2011
Online at stacks.iop.org/JNE/8/025028

Abstract
Advancing the brain–computer interface (BCI) towards practical applications in technology-based assistive solutions for people with disabilities requires coping with problems of accessibility and usability to increase user acceptance and satisfaction. The main objective of this study was to introduce a usability-oriented approach in the assessment of BCI technology development by focusing on evaluation of the user's subjective workload and satisfaction. The secondary aim was to compare two applications for a P300-based BCI. Eight healthy subjects were asked to use an assistive technology solution which integrates the P300-based BCI with commercially available software under two conditions—visual stimuli needed to evoke the P300 response were either overlaid onto the application's graphical user interface or presented on a separate screen. The two conditions were compared for effectiveness (level of performance), efficiency (subjective workload measured by means of NASA-TXL) and satisfaction of the user. Although no significant difference in usability could be detected between the two conditions, the methodology proved to be an effective tool to highlight weaknesses in the technical solution.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

An electroencephalographic-based brain–computer interface (BCI) represents a promising technology to provide severely motor-disabled people with an alternative channel for sending commands to the external world that does not rely on muscular activity (Wolpaw *et al* 2002, Kübler and Müller 2007).

To hold this promise EEG-based BCI technology needs to be moved from the laboratory to the home environment. In this context, evaluation of the system's usability is a mandatory step when designing and developing practical non-invasive BCI prototypes intended to be used in daily life.

According to ISO recommendations (ISO 9241 1998), the concept of usability can be split into three different measures: (i) effectiveness, which estimates the accuracy and the completeness with which intended goals are achieved;

[8] Author to whom any correspondence should be addressed.

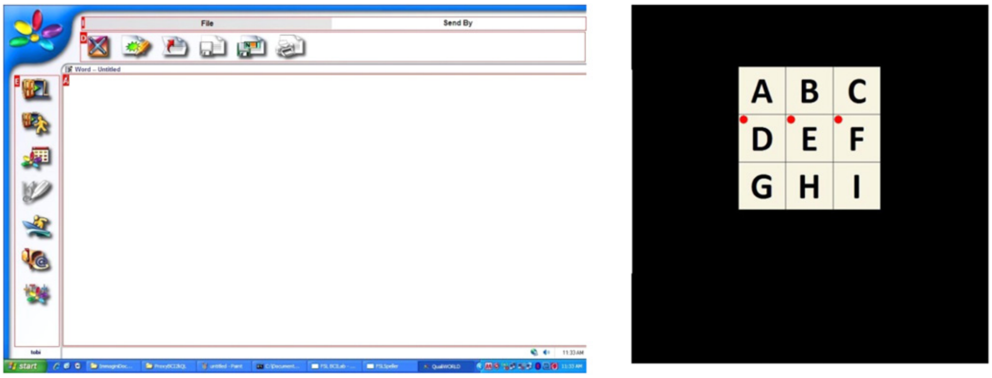

Figure 1. 'Split' condition. The first screen shows an assigned letter for each command available on the QW interface. The second screen shows the speller matrix on which the stimulation is presented.

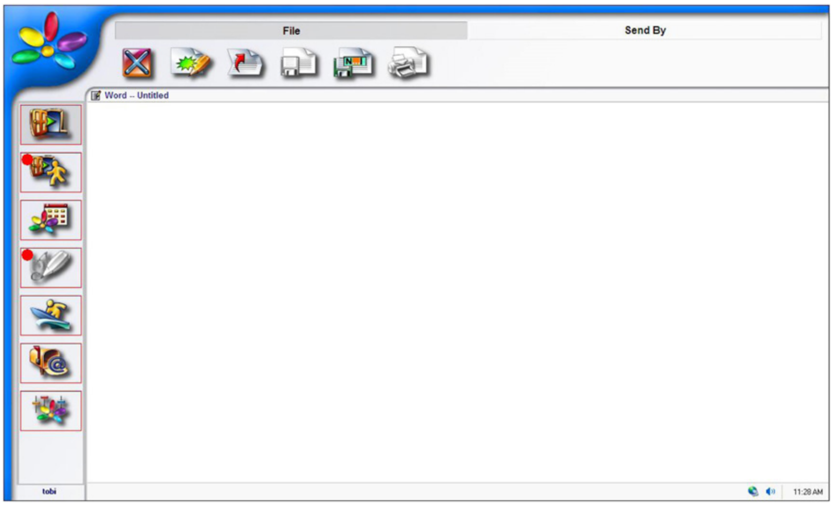

Figure 2. 'Overlaid' condition. The stimulation (red dots) is superimposed over the QW GUI. The QW commands are hierarchically organized. Selection of a command requires the user to first select the corresponding group of commands and then the command itself.

(ii) efficiency, which is the measure of the amount of human, economic and temporal resources expended in attaining the required level of product effectiveness; and (iii) satisfaction, a measure of the immediate and the long-term comfort and acceptability of the overall system.

The estimation of mental workload addresses the efficiency domain in terms of human resources and satisfaction. Hart and Staveland (1988) defined workload as a '*hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance*' (pp 230–50). Consequently, in this study we measured subjective workload and satisfaction to estimate the usability of a BCI system beyond the performance level.

In particular, this study was designed to evaluate the usability—derived from subject performance, workload and satisfaction—of an EEG-based BCI system integrated into the commercial accessibility software QualiWORLD (QW; QualiLife Inc. Paradiso-Lugano, CH) for communication and control (Riccio *et al* 2010). We used the event-related potential P300 as a BCI control signal because it also allows for high accuracy in patients with severe motor impairment (Kleih *et al* 2010, Nijboer *et al* 2008). Following the oddball paradigm (Duncan-Johnson and Donchin 1977) for eliciting the P300, visual stimulation was realized in two conditions: (i) two separate screens were presented to the subject, one on which the oddball was presented for the selection of commands (typical P300 matrix) and the other one with the QW application on which the commands were executed; (ii) subjects were presented with a single screen on which the oddball paradigm overlaid the application. In the latter condition, the users did not need to switch attention between two separate screens. This should reduce workload in mastering the BCI application.

2. Materials and methods

2.1. User interface

Figures 1 and 2 illustrate the user interface adopted in the two experimental conditions that will be referred to as 'split' and 'overlaid'.

In the '*split*' condition, similar to Mugler *et al* (2008), the computer with the QW-software was connected to two screens. The first screen displayed the QW graphical user interface (GUI) where the user could choose the desired command. To each of these commands, a letter was assigned

Table 1. Rating scale definitions of NASA-TLX factors.

Title	Endpoints	Descriptions
Mental demand	Very low/very high	How much mental and perceptual activity was required (e.g. thinking, deciding, remembering, looking, searching)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical demand	Very low/very high	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal demand	Very low/very high	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Perfect/Failure	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Very low/very high	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Very low/very high	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

that corresponded to a letter of the P300 matrix presented on the second screen. The users were asked to focus on the letter corresponding to the command they wanted to select (figure 1). We assigned red dots randomly flashing in rows and columns (as in a typical P300 matrix) to each letter to make stimulus presentation comparable to the "overlaid" condition.

In the '*overlaid*' condition, dots were assigned to each command of the QW GUI. These dots flashed randomly. No dedicated BCI window was visible to the users who, in this case, were asked to focus their attention directly on the desired control icon and to count mentally the number of times the dot appeared. The dot represented the target stimuli of the oddball paradigm and elicited a P300 potential (figure 2).

2.2. *Experimental procedure*

Eight healthy volunteers (six men, two women, 26.7 ± 1.3 years) agreed to participate in the study and provided written informed consent. The study protocol was approved by the local ethical committee. In both conditions, participants were asked to execute three different tasks: (i) to display an internet page in a browser and perform some browsing operations (*internet browsing*), (ii) to write a text in a word document (*word processing*) and (iii) to check QW settings and preferences, such as font size, mouse pointer, etc (*configuration of the software*). Both conditions were performed in a single session; the first task was randomly assigned to each subject.

The EEG recordings were performed with sintered Ag/AgCl electrodes mounted on a 16-channel cap (Electro-Cap International) and placed according to the 10–20 International System (Sharbrough *et al* 1991; Fp1, Fp2, F3, Fz, F4, T7, T8, C3, Cz, C4, CP3, CP4, P3, Pz, P4, and Oz). All 16 channels were referenced to linked earlobes. Impedances were kept below 5 kΩ. The EEG signal was amplified with commercial hardware (gUSBAmp amplifier, gTec, Austria), bandpass filtered between 0.1 and 60 Hz, and sampled at 256 Hz.

The BCI2000 (Schalk *et al* 2004) signal processing pipeline was used as a brain transducer. The fixed-length epochs following the onset of each stimulus were averaged on a stimulus class base. In this way a P300 ERP could be detected on the averaged waveform corresponding to the target stimulus. To each stimulus class, the classifier assigned a score using a set of signal features that was selected during the previous offline analysis phase. These features were computed using a step-wise linear discriminant analysis (SWLDA; Krusienski *et al* 2006) procedure.

A custom computer program was used to generate the overlaid visual stimulation, to initialize each trial with the correct number of targets (depending on the number of visible controls on the target application window) and to translate the BCI classification results into commands for the target application.

Effectiveness was measured in terms of the level of performance obtained under the two conditions ('overlaid' and 'split'). The level of performance was expressed as the ratio between the total time to successfully complete each task and the minimum number of selections needed to execute each single task.

Efficiency was estimated as subjective workload and satisfaction. The multidimensional NASA-TLX questionnaire (NASA Human Performance Research Group 1987) assesses subjective workload using six different factors (table 1). The subjective workload for each factor is rated on 20 step bipolar scales with a score from 0 to 100. The minimum and the maximum are anchored with phrases like 'very low' and 'very high' (table 1). A weighting procedure is used (a pairwise comparison task) to combine the six scale ratings into a total score (between 0 and 100). The questionnaire was self-completed by the subjects at the end of each task performed under both conditions, to capture the potential differences in workload. High scores at NASA-TLX are considered as an index of a high subjective workload associated with a particular task. High mental workload can affect the usability of the device.

User satisfaction was measured with a visual analogue scale (VAS) at the end of each condition. The subjects were asked to rate their 'overall satisfaction' with the BCI-device, drawing a vertical bar on a line where number 0 indicated that they were 'not satisfied at all', while number 10 meant that they were 'absolutely satisfied'.

The participant's preference between the two conditions was estimated by means of a bipolar VAS on which the

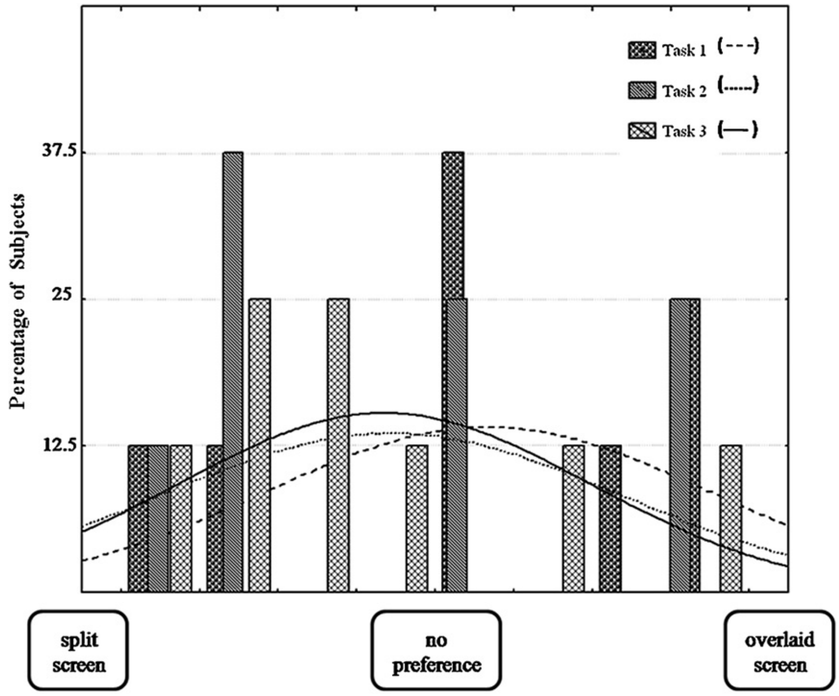

Figure 3. Distribution of subjects' preferences between 'split' and 'overlaid' conditions for each task. *y* axis: percentage of subjects expressing a given preference; *x* axis: type and level of perference between the two conditions. The fitting curve for each histogram is also shown.

'overlaid' and the 'split' condition represented the two poles and the centre represented 'no preference'. The subjects were asked to rate their preferences between the two conditions by drawing a vertical bar on a line where number 5 indicated 'no preference', while numbers from 0 to 4 and from 6 to 10 meant that they preferred either the 'split' or the 'overlaid' condition, respectively. The modality of scoring was not visible to the user. Two labels indicating the two different conditions were placed at the extremities of the VAS.

Finally, an unstructured interview was conducted with all subjects at the end of the session to assess their opinion about the two conditions and how difficult they found the two applications.

To reduce the variability of the data distribution within the sample ($n = 8$ subjects), a log-transformation was applied for statistical analysis. Two separate 2 (condition) $\times 3$ (task) repeated measure ANOVAs were calculated with the level of performance and the total NASA score as dependent variables. A *post hoc* Tukey test was chosen to assess significance ($\alpha \leqslant 0.05$). The student's *t*-test (paired) was performed to seek for significant differences between the subjects' level of satisfaction obtained under the 'overlaid' and 'spit' conditions.

3. Results

The mean values ($\pm$SD) of both the performance level and the NASA total scores corresponding to each task and both conditions for each subject are reported in table 2. No significant differences were found for the level of performance obtained for each of the three tasks, although a trend towards a longer time needed to perform tasks 1 (*internet browsing*) and 3 (*configuration of the setting)* was present under the 'overlaid' with respect to the 'split' condition.

For the workload total score, the ANOVA yielded a non-significant interaction between the task and the condition, but a significant main effect of the task ($F(2,14) = 6.15, p = 0.012$). *Post hoc* comparisons revealed task 2 (*word processing*) to require a lower workload than task 1 (*internet browsing*) ($p = 0.011$).

No significant differences between the satisfaction scores obtained for the two conditions ($p = 0.35$) were found.

The subjects' preference distribution is shown in figure 3. The fitting curves show that the subjects' preference for task 1 (internet browsing; dotted line) tends towards the 'overlaid screen' condition, whereas the preferences for task 2 (word processing; small dot line) and task 3 (configuration of the software; solid line) tend towards the 'split screen' condition.

During the unstructured interview, the subjects expressed their difficulties with the spatial presentation of the stimuli in a specific part of the screen (upper right corner of the GUI in figure 1) in the 'overlaid' condition. They reported the stimulation dots being very close to each other, especially for the internet browsing task. This is in line with the observed highest value of mental workload for this task in the 'overlaid' condition with respect to the others (see table 2).

4. Discussion and conclusion

Unlike we hypothesized, no significant differences were observed in efficiency (level of performance), user satisfaction and subjective workload measures between the 'overlaid' and

Table 2. Averaged performance values and NASA total score for different tasks and stimulation conditions.

	'Overlaid' condition tasks						'Split' condition tasks					
	Task 1		Task 2		Task 3		Task 1		Task 2		Task 3	
	LP[a]	WL[b]	LP	WL	LP	WL	LP	WL	LP	WL	LP	WL
Subject 1	35.0	74.7	30.9	66.3	24.1	70.7	25.9	86.0	46.7	80.0	24.1	73.3
Subject 2	44.8	80.7	34.8	71.7	24.1	85.0	23.4	86.7	58.5	56.7	26.0	86.7
Subject 3	39.7	38.0	49.0	43.7	24.1	40.0	33.5	57.3	37.7	40.3	20.8	48.0
Subject 4	55.1	84.0	62.6	64.3	78.4	42.3	32.1	53.7	45.4	36.3	34.6	46.0
Subject 5	60.0	58.3	51.2	25.3	60.0	53.7	60.0	36.0	50.5	14.7	60.0	15.7
Subject 6	36.8	76	30.9	60.7	39.5	77.7	66.5	59.0	53.5	46.7	33.9	59.3
Subject 7	60.0	43.7	62.6	15.3	31.5	9.0	48.7	9.3	30.9	29.0	41.5	10.0
Subject 8	60.0	75	44.3	69.3	60.0	66.7	27.3	73.0	30.9	75.0	26.0	70.0
Average	48.9	66.3	45.8	52.1	42.7	55.7	39.7	57.7	44.3	47.3	33.4	51.1
SD	11.0	17.5	12.9	21.5	20.8	24.8	16.5	25.9	10.2	22.3	12.7	27.2

[a] Level of performance (ratio between the time employed to complete the task and the minimum number of selections to complete it, expressed in seconds).
[b] Overall workload scores obtained applying NASA-TLX.
Task 1: internet browsing; task 2: word processing; task 3: configuration of the software.

'split' conditions. The only significance was related to the 'word processing' task (task 2) which was associated with a lower subjective workload with respect to the 'internet browsing' (task 1), irrespective of the condition. These latter findings can be ascribed to the similarity in the stimulus presentation based on a matrix for both the 'overlaid' and 'split' screens. Although this general lack of significance could be due to the sample size, other variables have to be taken into account.

In the overlaid condition, the subjective workload showed a tendency towards higher values for task 1 (browsing an internet page) with respect to those observed for the other two tasks (spelling a sentence and configuring the assistive application). A similar tendency was also observed for the level of performance (lower in the 'overlaid' condition). The physical position of the stimuli could have affected the level of performance of the overlaid interface. During a visual task, the reflexive orienting of visual attention, also known as an exogenous mechanism of attention (Jonides 1981), could drive the attention of the subject towards salient, task irrelevant, visual stimuli (Hickey *et al* 2006). In this perspective the exogenous attentional system could have shifted the subject's attention to the non-target stimuli close to the target stimuli during the internet task. This could have affected the performances of the subjects and increased the mental workload required to perform the task. The overall evaluation results prompted us to refine the 'overlaid' prototype and address a new testing study.

It is becoming evident that the importance of the subjective feedback from users and the usability evaluation are of the utmost relevance in the testing pipeline of a BCI system and its applications.

In this regard, the human factor methodologies, such as workload assessment using the NASA-TLX and user satisfaction measurement, together with accuracy measurement, are promising instruments for the evaluation of BCI-based prototypes. Usability studies can provide a well-suited approach for investigating human influences on the use of a BCI system. These analyses should be conducted during BCI system testing to adequately support the end users and to increase their acceptance of the system. In iterative processes, the BCIs can be adapted to the users so that in the future, BCIs for use in daily life will be available.

Acknowledgments

This work was supported by the European ICT Programme Project FP7-224631 (TOBI). The text reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein.

References

Duncan-Johnson C C and Donchin E 1977 On quantifying surprise: the variation of event-related potentials with subjective probability *Psychophysiology* **14** 456–67
Hart S G and Staveland L E 1988 Development of NASA-TLX (task load index): result of empirical and theoretical research *Human Mental Workload* ed P A Hancock and N Meshkati (Amsterdam: North Holland) pp 239–50
Hickey C, McDonald J J and Theeuwes J 2006 Electrophysiological evidence of the capture of visual attention *J. Cogn. Neurosci.* **18** 604–13
ISO 9241 1993 Ergonomic requirements for office work with visual display terminals (VDTs): part 11. Guidance on specifying and measuring usability International Standards Organization. Committee Draft
Jonides J P 1981 Voluntary versus automatic control over the mind's eye's movement *Attention and Performance IX* ed J B Long and A D Baddley (Hillsdale, NJ: Erlbaum) pp 187–203
Kleih S C, Nijboer F, Halder S and Kubler A 2010 Motivation modulates the P300 amplitude during brain–computer interface use *Clin. Neurophysiol.* **121** 1023–31
Krusienski D J, Sellers E W, Cabestaing F, Bayoudh S, McFarland D J, Vaughan T M and Wolpaw J R 2006 A comparison of classification techniques for the P300 speller *J. Neural Eng.* **3** 299–305
Kübler A and Müller K R 2007 An introduction to brain–computer interfacing *Toward Brain–Computer Interfacing* ed G Dornhege, R del J Millán, T Hinterberger, D J McFarland and K-R Müller (Cambridge, MA: MIT Press) pp 1–25

Mugler E, Bensch M, Halder S, Rosenstiel W, Bogdan M, Birbaumer N and Kübler A 2008 Control of an internet browser using the P300 event-related potential *Int. J. Bioelectromagn.* **10** 56–63

NASA Human Performance Research Group 1987 Task Load Index (NASA-TLX) NASA Ames Research Centre, NASA Human Performance Research Group http://humansystems.arc.nasa.gov/groups/TLX

Nijboer F *et al* 2008 A P300-based brain–computer interface for people with amyotrophic lateral sclerosis *Clin. Neurophysiol.* **119** 1909–16

Riccio A, Bianchi L, Aloise F, Zickler C, Hoogerwerf E J, Mattia D and Cincotti F 2010 Overlaid P300 stimulation-based BCI: as solution to 'limit' workload in communication application? *Proc. 4th Int. BCI Meeting* (Asilomar, California, USA, May 31–June 4 2010)

Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain–computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

Sharbrough F, Chatrian E G, Lesser P R, Luders H, Nuwer M and Picton T W 1991 American electroencephalographic society guidelines for standard electrode position nomenclature *J. Clin. Neurophysiol.* **8** 200–2

Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91