

Evaluating User Experience in a Selection Based Brain-Computer Interface Game

A Comparative Study

Hayrettin Gürkök, Gido Hakvoort and Mannes Poel

Human-Media Interaction, University of Twente, Enschede, The Netherlands
h.gurkok@utwente.nl, gido.hakvoort@gmail.com, m.poel@utwente.nl

Abstract. In human-computer interaction, it is important to offer the users correct modalities for particular tasks and situations. Unless the user has the suitable modality for a task, neither task performance nor user experience can be optimised. The aim of this study is to assess the appropriateness of using a steady-state visually evoked potential based brain-computer interface (BCI) for selection tasks in a computer game. In an experiment participants evaluated a BCI control and a comparable automatic speech recogniser (ASR) control in terms of workload, usability and engagement. The results showed that although BCI was a satisfactory modality in completing selection tasks, its use in our game was not engaging for the player. In our particular setup, ASR control appeared to be a better alternative to BCI control.

Keywords: Brain-computer interface, steady-state visually evoked potential, games, user experience, engagement, usability, workload.

1 Introduction

Traditional human-computer interaction (HCI) modalities, mouse and keyboard, have long served as a reliable means of input. Despite their reliability, they restrict users expressive capabilities and therefore the information transferred from the user to the computer. As a response to this problem, modern HCI uses natural human input such as speech and gestures. Moreover, through brain-computer interfaces (BCIs), even brain activity can be provided as input to computers.

In HCI, it is important to offer the correct modalities for specific tasks and situations. BCIs, for example, have long been in use for selection tasks in assistive [14] and also recreational [15] applications. However, in all these applications, the primary concern has been to optimise the recognition performance so the appropriateness and consequences of using BCIs for selection tasks has never been considered. In entertainment computing, the primary goal of the user is not to optimise task performance. While playing a game, the user (i.e. player) might still have tasks to complete but their actual purpose in playing the game is enjoyment. They may attempt to fulfil the tasks in the game or may simply wander in the virtual environment (VE). They may get bored using the mouse and the keyboard, which

offer no challenge to them, but enjoy tackling the shortcomings of a non-traditional modality, such as BCI [12]. Especially for BCIs, which provide relatively low throughput [17], factors such as usability and workload should carefully be taken into account. Unless the user has the suitable modality for a task, neither task performance nor user experience can be optimised.

The aim of this study is to assess the appropriateness of using a steady-state visually evoked potential (SSVEP) based BCI for selection tasks in a computer game. We evaluate BCI control subjectively in terms of user experience related factors, which are workload, usability and engagement, using questionnaires. We contrast the evaluation results for BCI control with control with a comparable input modality, an automatic speech recogniser (ASR), in the same game. BCI and ASR are both imperfect but natural input modalities. Moreover, they are suitable candidates for multimodal control in gaming applications, considering the assumption that the majority of the primary game controllers already occupy the players' hands. In this study, ASR control is a reference for assessing BCI control.

2 Background

2.1 Evaluation of User Experience in BCIs

In evaluating BCI applications, the traditional concern is improving the recognition performance. Nevertheless, user experience has recently been considered in BCI system evaluations. The most commonly evaluated aspect is the workload imposed by the BCI. Frequently the NASA Task Load Index (NASA-TLX) [5] is used in evaluating BCI systems based on selection tasks. Example studies include the evaluation of spellers [13] and other assistive communication applications [16, 19]. In one of these examples [13] the usability of a speller was also evaluated using the System Usability Scale (SUS) [3]. In another one [16] overall user satisfaction was measured in a communication application by using a 10-point visual analogue scale. Subjective assessment of presence in a VE controlled by BCI has also been demonstrated [4].

As the aforementioned studies indicate, subjective evaluation is a commonly practised and suitable technique for evaluating BCI systems. The questionnaires are especially easy and comfortable to apply, suitable for extracting statistical analyses quickly, strong and reliable once validated, and applicable to the majority of BCI users [9]. Note that all the studies mentioned so far in this subsection involve assistive BCI systems, targeting disabled people. Studies evaluating user experience in BCI entertainment applications are extremely rare. An example study [10] suggested a user experience questionnaire for BCI games although the game used in the study was a simple, controlled game. To our knowledge, no work has been done on systematic evaluation of user experience in realistic BCI games. Hence the time has come to set up user experience evaluation tools for these games.

2.2 BCI, SSVEP and BCI Games

BCIs can infer a user's mental/emotional state or intention by interpreting brain activity and use this information to control computer applications. First, brain activity is acquired and quantified as a signal, which is mostly done through the use of an electroencephalograph (EEG). EEG measures electrical brain activity via electrodes in contact with the scalp. Then, the signal is processed and analysed using neuromechanisms. Neuromechanisms signify certain changes in the signal with respect to an event. The event can be a voluntary action such as moving a hand or looking at something as well as an involuntary reaction to a stimulus or an error. Finally, according to the result of signal analysis, a control signal is sent to an application.

SSVEP is a stimulus dependent, widely used neuromechanism. When a person attends to a visual stimulus repeating with a certain frequency, the amplitude of the signal measured from the visual brain area is enhanced at the frequency of the stimulation. This enhancement is known as the SSVEP [8]. SSVEP is frequently used for selection tasks. By presenting multiple stimuli with distinct repetition frequencies, it is possible to detect which of the stimuli a person was paying attention to. So if each of these stimuli is associated with a choice, then it is possible to detect the person's selection. The strength of the SSVEP is dependent on the stimulation properties. These include, but are not limited to, flicker frequency, size, colour and shape of the stimulus [1].

We can identify two genres of BCI games: active and passive BCI games. In the former, the player intentionally tries to regulate their brain activity while in the latter the game captures naturally-occurring brain activity. So, a game in which the player concentrates on a flickering image to produce an SSVEP would be an active BCI game while a game which changes the speed of an avatar according to player's state of relaxedness would be a passive BCI game. The principles of active and passive BCI games are analogous to directly and indirectly controlled physiological games. Recent research has shown that direct physiological control affords a better user experience due to its visible responsiveness [11]. We also opted for direct control, in other words for an active BCI game, while developing the game used in this study.

3 Method

3.1 Rationale

With BCIs, especially with the stimulus-dependent ones such as the SSVEP-based BCI which we also used in our study, many factors deserve attention while evaluating gaming experience. The flow provided by the game, the workload imposed by the stimulation, player's comfort and safety in relation to the stimulation, usability and intuitiveness of the interaction need separate assessment. At the moment, there is no standard questionnaire or method to evaluate BCIs collectively for user experience factors. This is why, as we mentioned in section 2.1, multiple questionnaires are used together for evaluating BCIs.

To understand whether SSVEP based BCI is a suitable modality for selection tasks, we assessed user experience during BCI control in a computer game with respect to three concepts, namely workload, engagement and usability. We use standard questionnaires for assessing these concepts. We also compare the results against ASR control in the same game, as a reference condition. So, the analysis results can shed light on the appropriateness of using BCI as well as ASR for selection tasks. To support the interpretation of subjective findings, we analyse the objective data corresponding to player performance and effort, both of which may have a role in user experience.

3.2 The Game: Mind the Sheep!

Mind the Sheep! (see Fig. 1) is a multimodal computer game where the player needs to herd a flock of sheep across a field by commanding a group of dogs. The game world contains three dogs, ten sheep, a pen and some obstacles. The aim is to pen all the sheep as quickly as possible. For the purpose of this work, we used the BCI and ASR controlled versions of the game.



Fig. 1. Screenshots from the game. From left to right: BCI game with stimulation off, BCI game with stimulation on, and ASR game.

To command a dog, the player positions the cursor at the point to which the dog is supposed to move. The player holds the mouse button pressed to provide the command to select the dog. Meanwhile, the game displays cues specific to the active modality (ASR or BCI). When ASR is the active modality, names appear under the dog images and the player pronounces the name of the dog they want to select. When BCI is the active modality, dog images are replaced by circles flickering at different frequencies and the player concentrates on the circle replacing the dog they want to select (so as to obtain an SSVEP). The stimulation persists and, depending on the active modality, EEG or acoustic data is accumulated as long as the mouse button is held. When the player releases the mouse button, the signal is analysed and a dog is selected based on this analysis. The selected dog immediately moves to the location where the cursor was located at the time of mouse button release.

3.3 Questionnaires

We use the NASA-TLX [5] for workload evaluation as it is a brief, yet powerful questionnaire used frequently in BCI research. The NASA-TLX measures subjective workload for a task using six items which assess mental demand, physical demand, temporal demand, performance, effort and frustration. Each item is rated using a 20-step bipolar scale resulting in a score between 0 and 100. The low and high ends of each scale are anchored with a word pair indicating the two extremes for the item (e.g. word pair perfect-failure for performance). An average or a weighted average of item scores provide the overall workload score. A higher score implies a higher subjective workload associated with a task. In our study we used an unweighted version of the NASA-TLX.

For engagement assessment we used the Game Engagement Questionnaire (GEQ) [2] because, to our knowledge, it is the only well-defined and validated questionnaire designed specifically for games. The GEQ is a questionnaire measuring subjective level of game engagement in four dimensions which are absorption, flow, presence and immersion. Nineteen items, each formed as a statement, are marked on a 5-point bipolar scale with respect to the level of agreement. Columns corresponding to the low end, middle and high end points of the scale were anchored with words No, Maybe and Yes respectively. The points are averaged over the items to reach the overall engagement score. A higher score indicates a higher level of engagement in the game.

Our choice of questionnaire for assessing usability is based on a previous comparative study [18]. In that study, AttrakDiff2 [6] was suggested as the most suitable usability evaluation method for a broad range of multimodal applications. AttrakDiff2 has a reduced version as well [7], which we also use in this study, making it more convenient to fill in and analyse. The AttrakDiff2 evaluates product quality in three dimensions which are pragmatic quality, hedonic quality and attractiveness. Pragmatic attributes are those relating to a product's functionality (i.e. utility) and ways to access functionality (i.e. usability). Hedonic attributes, on the other hand, are those that provide stimulation and communicate user identity. The questionnaire contains twenty-one items rated using a 7-point semantic differential scale. For each item, the scale is anchored at extremes by opposite word pairs (e.g. simple-complicated). Ratings averaged over the items imply an overall product quality score. The higher the score, the better the subjective quality of the product. Although our aim is to evaluate the usability (i.e. pragmatic quality) of the interface, we kept the items for the other two dimensions in as well.

4 Experiment

4.1 Participants

Twenty people (3 female) participated in the experiment. They had an average age of 24.9 ($\sigma = 2.87$), ranging from 19 to 29 years, and normal or corrected vision. None of them were native English speakers. Eight of them had previous experience with BCIs

and fourteen of them with ASRs. Six of them indicated that they played games more than five hours per week. Informed consent was obtained from all participants and they were paid according to the regulations of our institution.

4.2 Game Parameters

Ensuring the equivalence of the ASR and the BCI in terms of recognition performance was a concern, as this could highly affect the game experience. We did not want to artificially deteriorate the performance of modalities by introducing noise or random errors but we did try to equalise the performances by tuning game parameters.

We conducted two pilot studies to standardise the recognition performances of the ASR and the BCI. For the ASR, we tested for different sets of dog names; and for the BCI, we tested for different sets of frequencies and sizes for the flickering circles. We decided to use Dexter, Lassie and Shadow as dog names, 7.5 Hz, 10 Hz and 12 Hz as flicker frequencies and 3 cm as the flicker diameter length. Literature also confirms that flicker frequencies between 5-12 Hz can evoke strong SSVEP and size of 3 cm can provide an optimal comfort-performance combination [1].

4.3 Procedure

Participants sat on a comfortable chair approximately 60 cm away from a 20" screen with a resolution of 1280 × 960. They played Mind the Sheep! two times in total; once with BCI and once with ASR in counterbalanced order. They played each game until all the sheep were penned or the play time reached 10 minutes. After each game, they filled in the three questionnaires, NASA-TLX, GEQ and AttrakDiff2, in the given order. In the NASA-TLX, the "task" was defined as "selecting a dog". For the AttrakDiff2, the "product" was replaced with "the interface for commanding the dogs" and participants were instructed to complete the questionnaire with respect to the devices they would need to use and tasks they would need to perform to select a dog.

In the ASR game, BCI control was not available and brain signals were not analysed. Sound was acquired by the microphone located to the right, behind of the participants. This particular location was chosen in order to match the ASR recognition performance with that of the BCI, as described in the previous subsection. In the BCI game, ASR was not available and speech was not recognised. Brain signals were acquired by five EEG electrodes placed on participant's head. During all games, each key press and mouse click was logged along with a timestamp. The game world layout was different in each game but comparable in difficulty.

4.4 Analysis

For workload and engagement analysis, we computed the means of NASA-TLX and GEQ scores respectively over all participants. For usability analysis, we did the same

only with the pragmatic quality scores of the AttrakDiff2. In NASA-TLX we also computed the mean score per item and in GEQ the mean score per dimension. Furthermore, we analysed log data to support our interpretation of the questionnaire results. We computed average selection durations as an indicator of effort and total number of selections (i.e. number of times the mouse button was released) and total game durations as indicators of performance.

Although we computed and report the means in analysis results, we opted for non-parametric statistical testing for assessing the significance of all differences since we neither can assume nor could prove normally distributed samples. Thus, the significant differences mentioned throughout the next subsection were assessed by the Wilcoxon signed-rank test ($p < 0.05$). Unless otherwise stated, reader should assume non-significant difference.

4.5 Results

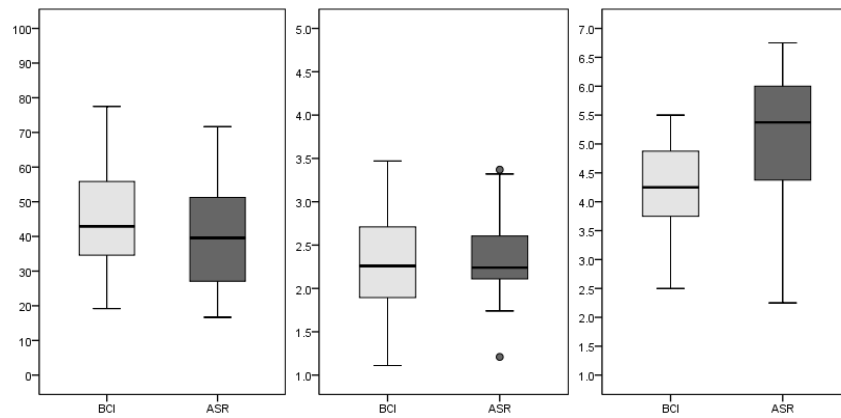


Fig. 2. Box plots of questionnaire scores for both games. From left to right: NASA-TLX, GEQ and AttrakDiff2 pragmatic quality scores corresponding to workload, engagement and usability levels respectively. The difference in AttrakDiff2 scores is significant.

Fig. 2 displays the box plots summarising the three questionnaire scores for both games. NASA-TLX (i.e. workload) scores were higher for the BCI game ($\mu = 45.21$, $\sigma = 15.12$) than those for the ASR game ($\mu = 40.08$, $\sigma = 14.85$) and AttrakDiff2 pragmatic quality (i.e. usability) scores were higher for the ASR game ($\mu = 5.08$, $\sigma = 1.27$) than those for the BCI game ($\mu = 4.25$, $\sigma = 0.85$). GEQ (i.e. engagement) scores were comparable for BCI ($\mu = 2.31$, $\sigma = 0.62$) and ASR ($\mu = 2.35$, $\sigma = 0.53$) games. The difference in AttrakDiff2 pragmatic quality scores was significant.

There was no significant difference between the two games in any GEQ engagement dimension score (see Table 1). A deeper analysis, at item level, did not yield any significant difference either.

Table 1. Mean and standard deviation (in parentheses) values of scores for the GEQ engagement dimensions.

Dimension	BCI	ASR
Presence	2.51 (0.70)	2.45 (0.67)
Absorption	1.88 (0.77)	1.79 (0.57)
Flow	2.38 (0.69)	2.51 (0.65)
Immersion	3.10 (1.25)	3.25 (1.25)

Item level analysis of the NASA-TLX workload questionnaire revealed that the mental demand item was significantly different between the two games (see Table 2). There was also a trend for significant difference in the effort item ($p = 0.076$).

Table 2. Mean and standard deviation (in parentheses) values of scores for the NASA-TLX workload items. (* significant difference, † trend toward significant difference with $p = 0.076$)

Item	BCI	ASR
Mental Demand*	52.00 (21.67)	39.50 (24.06)
Physical Demand	30.50 (21.39)	33.50 (21.40)
Temporal Demand	49.25 (25.56)	46.50 (19.13)
Performance	31.50 (17.93)	32.00 (22.68)
Effort†	58.78 (20.64)	49.25 (20.28)
Frustration	49.25 (27.40)	39.75 (24.79)

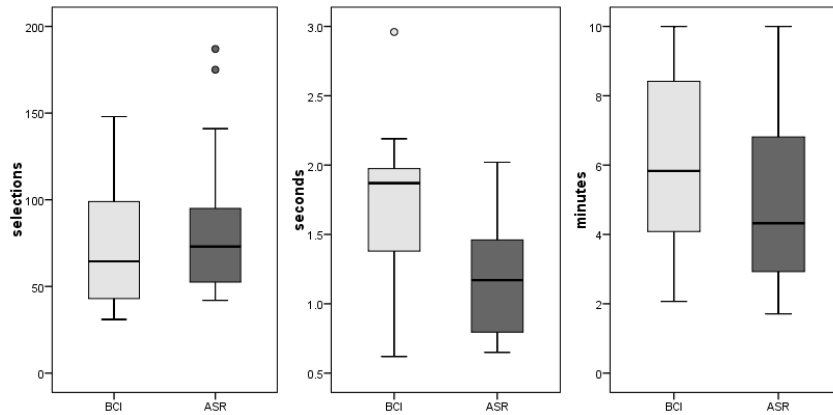


Fig. 3. Box plots for log analysis results. From left to right: number of selections, average selection duration, and game duration. The differences in the latter two are significant.

Fig. 3 displays the results of the log analysis in box plots. The number of selections was less during the BCI ($\mu = 72.35$, $\sigma = 33.05$) game than that during the ASR ($\mu = 84.25$, $\sigma = 41.28$) game. Average selection duration (in seconds) was significantly longer in the BCI game ($\mu = 1.71$, $\sigma = 0.53$) than that in the ASR game ($\mu = 1.17$, $\sigma =$

0.37). Furthermore, total game duration (in minutes) was also significantly longer for the BCI game ($\mu = 6.29, \sigma = 2.47$) than that for the ASR game ($\mu = 4.98, \sigma = 2.63$).

5 Discussion

5.1 Workload and Performance

The results of the NASA-TLX workload analysis showed that participants experienced slightly higher workload during the BCI game in comparison to the ASR game. But neither of the games imposed high workload, as the mean scores for both were below the median value (50/100). Lower level analysis revealed that the difference in the performance item was negligible between the two games. Since in the NASA-TLX we asked the questions for the task of selecting dogs (thus not for the game as a whole), this way, our pilot study to equalise the recognition performances of the two modalities was validated subjectively. Also objectively, the log analysis showed that the number of selection attempts were comparable in both games. This suggests that participants achieved similar performances for selecting dogs in both games. On the other hand, it took significantly more time for the participants to complete the BCI game than the ASR game. This might suggest that ASR performed better than BCI. However we note that the game completion time is not dependent only on the recognition performance of the modalities but also on modality specific strategies. For example, as the log analysis showed, the average duration of selections was significantly higher in the BCI game than that in the ASR game. In this case, while making a selection in the BCI game, the game state would change more than it would in the ASR game. This might necessitate recreating a strategy after some selections in the BCI game thus increase the game completion time.

Mental workload was the item which differed the most significantly between the two games. It was followed, though non-significantly, by effort and frustration. Altogether, these imply that although the participants could achieve comparable performance during both games, the task of selecting dogs was more mentally demanding in the BCI game so that they had to put more effort in. Achieving the same end result by putting in more effort might have introduced more frustration during the BCI game.

With respect to temporal and physical demand, there were no significant differences. There was no deadline for completing a task in both games so the temporal demand, which is the pace of the task, was independent of the control modality. Therefore the absence of difference in temporal demand is in line with the game logic. We understand that looking at the stimulation in the BCI game and speaking in the ASR game were equally demanding. Here we note that during the ASR game some participants needed to speak louder than they would do in real life since the microphone was located behind them for the sake of equalising modality recognition performances.

5.2 Engagement

There was no difference in the overall GEQ engagement scores for the BCI and ASR games. There was also no difference in any questionnaire dimension or item. This implies that modality was not a factor in engagement during the game. The total engagement scores for both games were lower than the median (2.5/5.0) so we understand that the game itself might not be engaging enough. Participants did not confirm that the game was not engaging but some of them indicated that the GEQ seemed more suitable for more immersive games.

5.3 Usability

AttrakDiff2 pragmatic quality scores for both modalities were above the median (3.5/7.0) so both inputs were satisfactory in terms of usability. Nevertheless, the scores were significantly higher for the ASR. Thus, speech input was rated higher than thought input for the usability aspects covered in the questionnaire such as simplicity, predictability and practicality. Note that, our instructions for AttrakDiff2 instructed participants to consider the devices they used for selections. In the ASR game, participants had no contact with the microphone. In the BCI game we placed a cap on participants' head, applied some electrolyte gel in their hair, and connected the electrodes placed on their head to an immobile EEG. Consequently, the additional setup time and lack of freedom to move might have decreased the usability of BCI.

5.4 Assessment and Limitations of the Study

Although we used validated questionnaires in this study, their effectiveness is highly influenced by the appropriateness of the context they are used in. NASA-TLX is a brief questionnaire, used regularly in BCI studies as we also said in our survey in section 2.1. In our study we also obtained invaluable insight into workload using this questionnaire. AttrakDiff2 is often used for evaluating the usability of commercial products. The role of usability in gameplay might seem to be negligible but, as we have also showed in this study, it can make a difference when comparing modalities in a game environment. Despite the fact that GEQ was developed for measuring engagement in computer games, in our study it did not prove to be effective. Participants had problems especially in answering the negated questionnaire items and understanding some terms. Thus, it is doubtful whether the information we obtained out of GEQ is reliable.

This work is not a conclusive study on the use of BCI in computer games. We used a particular neuromechanism, SSVEP, which is very frequently used in BCI applications, especially for selection tasks. Further research is necessary with other neuromechanisms, tasks and types of games. Perhaps a neuromechanism independent of stimulation can impose much less workload. For example imagining movements or trying to relax are commonly used actions in BCI applications and do not require any stimulation [15]. Maybe a more intuitive way of commanding dogs can provide better usability. We usually look at things when we want to interact with them but they do

not flicker in real life. Whereas calling a dog by its name is quite intuitive. This is a possible reason that might explain the higher usability during the ASR game in our study. The game we used had a restricted 2D environment with few agents. Therefore it might not be the best platform to do an engagement study. A more complex game, perhaps modification of a popular casual game so as to include BCI, might be more suitable to conduct engagement research. The EEG we used to acquire brain activity was an immobile device restricting bodily movements to some extent and requiring application of electrolyte gel in the hair. Using a portable EEG device with dry electrodes might improve the usability during the BCI game.

6 Conclusions

The purpose of this work was to evaluate user experience during SSVEP based BCI control in order to assess whether it is a suitable modality for selection tasks in HCI. We considered subjective evaluation of three factors related to user experience, which are workload, engagement and usability, while playing a BCI computer game. The evaluations are compared against ASR control in the very same game which can be considered as a reference condition.

BCI control imposed a workload below the median and provided usability above the median of the questionnaire scales we used. On the other hand, it yielded a level of engagement that is less than the median of the used scale. These suggest that although BCI might be a satisfactory modality in completing selection tasks, its use in games for selections might not be entertaining for the player. ASR control gave a similar overall impression except that it was significantly easier to use. BCI scored above the average with respect to mental demand and effort while ASR scored below the average. But the perceived performances of the modalities were not different. Achieving the same end result by putting in more effort might have introduced the higher frustration found for the BCI game. All these findings suggest that although BCI is a suitable modality for selection tasks, there is at least one better alternative, ASR.

Acknowledgments. The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. They would also like to thank L.E. Packwood for her help in improving the language of the article.

References

1. Bieger, J., Molina, G.G.: Light stimulation properties to influence brain activity: A brain-computer interface application. Tech. Rep. TN-2010-00315, Philips Research, Eindhoven, The Netherlands (2010)
2. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology* 45(4), 624-634 (2009)

3. Brooke, J.: SUS: a 'quick and dirty' usability scale. In: *Usability Evaluation in Industry*, pp. 189-194. Taylor & Francis Ltd., London, UK (1996)
4. Groenegrass, C., Holzner, C., Guger, C., Slater, M.: Effects of P300-based BCI use on reported presence in a virtual environment. *Presence: Teleoperators and Virtual Environments* 19(1), 1-11 (2010)
5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Human Mental Workload*, pp. 139-183. North-Holland, Amsterdam, The Netherlands (1988)
6. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: *Mensch & Computer 2003*, pp. 187-196. B.G. Teubner, Stuttgart, Germany (2003)
7. Hassenzahl, M., Monk, A.: The inference of perceived usability from beauty. *Human-Computer Interaction* 25(3), 235-260 (2010)
8. Herrmann, C.S.: Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research* 137(3-4), 346-353 (2001)
9. van de Laar, B., Gürkök, H., Plass-Oude Bos, D., Nijboer, F., Nijholt, A.: Perspectives on user experience evaluation of brain-computer interfaces. In: *Universal Access in Human-Computer Interaction. Users Diversity.*, pp. 600-609. Springer-Verlag, Berlin/Heidelberg, Germany (2011)
10. van de Laar, B., Reuderink, B., Plass-Oude Bos, D., Heylen, D.: Evaluating user experience of actual and imagined movement in BCI gaming. *International Journal of Gaming and Computer-Mediated Simulations* 2(4), 33-47 (2010)
11. Nacke, L.E., Kalyn, M., Lough, C., Mandryk, R.L.: Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In: *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*. pp. 103-112. ACM, New York, NY, USA (2011)
12. Nijholt, A., Reuderink, B., Oude Bos, D.: Turning shortcomings into challenges: Brain-computer interfaces for games. In: *Intelligent Technologies for Interactive Entertainment*, pp. 153-168. Springer-Verlag, Berlin/Heidelberg, Germany (2009)
13. Pasqualotto, E., Simonetta, A., Gnisci, V., Federici, S., Belardinelli, M.O.: Toward a usability evaluation of BCIs. *International Journal of Bioelectromagnetism* (2011), to appear.
14. Pfurtscheller, G., Müller-Putz, G., Scherer, R., Neuper, C.: Rehabilitation with brain-computer interface systems. *Computer* 41(10), 58-65 (2008)
15. Plass-Oude Bos, D., Reuderink, B., van de Laar, B., Gürkök, H., Mühl, C., Poel, M., Nijholt, A., Heylen, D.: Brain-computer interfacing and games. In: *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*, pp.149-178. Springer-Verlag, London, UK (2010)
16. Riccio, A., Leotta, F., Bianchi, L., Aloise, F., Zickler, C., Hoogerwerf, E.J., Kübler, A., Mattia, D., Cincotti, F.: Workload measurement in a communication application operated through a P300-based brain-computer interface. *Journal of Neural Engineering* 8(2), 025028 (2011)
17. Tonet, O., Marinelli, M., Citi, L., Rossini, P.M., Rossini, L., Megali, G., Dario, P.: Defining brain-machine interface applications by matching interface performance with device requirements. *Journal of Neuroscience Methods* 167(1), 91-104 (2008)
18. Wechsung, I., Naumann, A.: Evaluation methods for multimodal systems: A comparison of standardized usability questionnaires. In: *Perception in Multimodal Dialogue Systems*, pp. 276-284. Springer-Verlag, Berlin/Heidelberg, Germany (2008)
19. Zander, T.O., Gaertner, M., Kothe, C., Vilimek, R.: Combining eye gaze input with a brain-computer interface for touchless human-computer interaction. *International Journal of Human-Computer Interaction* 27(1), 38-51 (2011)