

Perspectives on User Experience Evaluation of Brain-Computer Interfaces

Bram van de Laar, Hayrettin Gürkök, Danny Plass-Oude Bos,
Femke Nijboer, and Anton Nijholt

Human Media Interaction, University of Twente, Enschede, The Netherlands
{b.l.a.vandelaar,h.gurkok,d.plass,f.nijboer,a.nijholt}@cs.utwente.nl

Abstract. The research on brain-computer interfaces (BCIs) is pushing hard to bring technologies out of the lab and into society and onto the market. The nascent merge between the field of BCI and human-computer interaction (HCI) is paving the way for new applications such as BCI-controlled gaming. The evaluation or success of BCI technologies is often based on how accurate the control of a user is with the technology. However, while this is still key to its usability, other factors that influence the user experience (UX) can make or break a technology. In this paper we first review studies which investigated user experience with BCIs. Second, we will discuss how HCI approaches can contribute to the evaluation of BCIs. Finally, we propose to develop a standardized questionnaire for evaluating BCIs for entertainment purposes.

1 Introduction

Brain-computer interfaces (BCIs) traditionally aimed to provide a reliable control signal for assistive technology for disabled persons. With the merge of the fields of human-computer interaction (HCI) and BCI new applications are being developed for entertainment and education which may be interesting for disabled *and* abled users. The aim of such applications is to create positive experiences that enrich our lives rather than only provide reliable control. To evaluate such systems, the user experience (UX) needs to be the central focus of study, rather than the reliability of the control signal, so that we can better understand how such a system can satisfy the needs of the user. Even though user experience evaluation is not common in current BCI studies, the user's experience can influence objective performance measures, such as BCI classifier accuracies, and has a big impact on whether users are actually willing to use a specific system.

In this paper, we review studies that investigate user experience in BCI research and the benefits of including such evaluations. Then, we will argue how the use of various techniques from the field of HCI can be advantageous for evaluating BCIs. In the last part of this paper we will elaborate on our efforts to develop a standardized questionnaire for evaluating user experience with BCIs.

2 Current State of User Experience Evaluation of BCI

2.1 User Experience Affects BCI

The BCI studies that do include UX indicate three main reasons to evaluate the UX: to increase user acceptance, to improve performance of the system, and to increase enjoyment.

User-centered approaches can increase usability and user acceptance, which is why some BCI groups involve users in the design process. They assess user needs and develop user requirements [1,2]. Pasqualotto et al. compared the usability and workload of two systems using the System Usability Scale (SUS) and NASA Task Load Index (TLX) [3,4,5] Developing user requirements is only the first step. The next is to assess the UX and user acceptance in a structured way *during* or *directly after* a user has interacted with the system. Subjective and experiential factors can shed light on the UX, combined with objective usability measures.

Several BCI studies suggest a relation between motivation and BCI task performance [6,7], and small but significant effects have been found [8] using an adapted version of the Current Motivation Questionnaire, a questionnaire for current motivation in learning and performance situations [7,9]. Similarly, the users' belief of how accurately they can control a BCI has an influence on their actual performance. Barbero and Grosse-Wentrup observed that participants who normally perform around chance level, perform better when they think they are doing better than they actually are (positive bias). Capable participants, however, performed worse when given inaccurate feedback, whether the bias was positive or negative [10]. Motivation may be only one of the performance-related factors that are influenced by the UX. By evaluating and improving the UX, other relations between the user and BCI recognition performance could be exploited to improve performance measures. There could also be mechanisms with indirect influences. For example, a system that is perceived as more beautiful is also perceived as more usable [11]. This perception could influence motivation which in turn could influence performance.

Most BCI applications currently being developed still serve only as a proof of concept [12], which may be why the entertainment value is often not evaluated. An exception is the BCI game BrainBasher, which was evaluated for the influence of different graphical interfaces and different user tasks, using the Game Experience Questionnaire which addresses immersion, tension, competence, flow, negative affect, positive affect, and challenge [13,14,15]. In the first study, the UX and performance have been determined for a clinical setup with minimal information on the screen. This was compared to a game-like setup of exactly the same task. The game version resulted in higher immersion. The second study compared the UX for imaginary and actual movement. Imagined movement was perceived as more challenging, but when using actual movement the participants stayed more alert.

Although there is not much research yet in this area, important questions arise, and the studies that have been done make clear that the UX can affect a BCI system in important ways.

2.2 BCI Affects User Experience

UX can influence the performance of BCIs, but BCIs can affect the UX as well, in two ways: (1) by using information about the user's mental state to adapt the interface or the interaction itself, with as goal to improve usability and UX; and (2) through the effects of using this particular input modality.

In the first approach, the user is supported in the tasks they are trying to accomplish with the help of BCI, which in turn could increase user satisfaction. For example, error-related brain activity can be detected and used to fix user or system errors for improved error handling [16]. The amount of information presented on screen can be adjusted according to the user's workload [17]. BCI could also be used to create or maintain specific experiences. As an example, brain activity indicators of stress or boredom can be used to keep the user in the optimal state of flow [18].

On the other hand, using BCI for input can already influence the UX by itself. Friedman et al. investigated whether the use of imaginary movement to walk in a virtual world would increase presence, using the Slater-Usoh- Steed presence questionnaire combined with a non-structured interview [19,20]. In a follow-up experiment, Groenegress et al. compared the presence experienced with a P300 interface to eye gaze and wand navigation [21]. Both experiments concluded that the BCI did not have a positive influence. However, the influence of BCI on UX is not limited to presence alone. Besides, BCI inputs are not limited to these two mental tasks. If the influence of certain mental tasks in specific contexts would be known, it would be possible to find the optimal mental tasks for a given application.

Whether BCI is used to affect the UX purposefully or whether this happens just by using this input modality, in both cases it is important to evaluate and be aware of the effects.

3 Applying HCI User Experience Evaluation to BCIs

Although evaluating the usability and UX of BCI systems is not common practice, in HCI research and development, especially for entertainment technologies which simply aim to improve the well-being of users, UX is a major concern. Therefore, the HCI community designs for UX and develops methods to evaluate it. Current methods for evaluating UX in entertainment technologies can be classified into quadrants of a plane which has an objective versus subjective axis and a qualitative versus quantitative axis [22] (see Fig. 1). The objective methods are based on overt and covert user responses during interaction while the subjective methods rely on user expressions after the interaction. The quantitative methods employ statistical analysis on collected data whereas the qualitative methods are interpretations of user responses by researchers. Below, we describe the methods corresponding to the quadrants formed by these two axes and discuss their contribution in evaluating BCI systems.

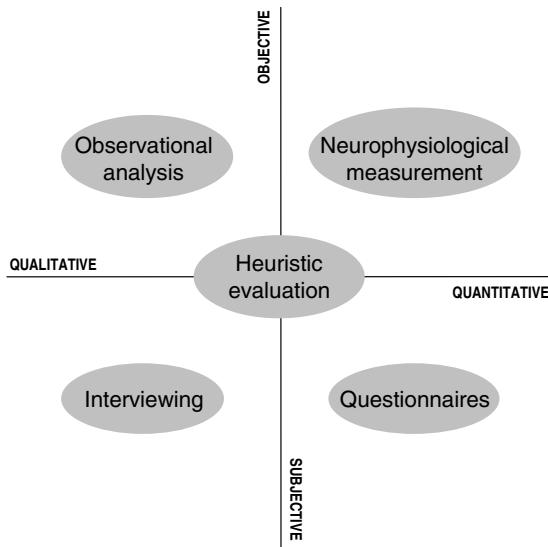


Fig. 1. A classification of current user experience evaluation methods used in human-computer interaction for entertainment technologies (adapted from [22])

3.1 Observational Analysis

Observational analysis is a qualitative-objective method as it relies on interpretation of overt user response by researchers. The classical way of observing overt user behaviour is through audiovisual recorders which provide qualitative data for gestures, facial expressions and verbalisations. There are some difficulties associated with annotating and analysing such rich data though. Firstly, while analysing the data, the researchers should acknowledge their biases, address inter-rater reliability and not read inferences where none are present. Secondly, there is an enormous time commitment associated with observational analysis. The ratio of analysis time to data sequence time ranges from 5:1 to 100:1 [23]. Thirdly, the operation of audiovisual recorders impose restrictions such as noise-free environment during audio recording or consistent illumination during video capturing. Similar restrictions are also imposed by brain activity recording devices. For example, the electroencephalogram (EEG, measuring electrical brain activity) is affected by user's movement [24], so users are usually asked to keep their body and face motionless. Thus, overt behaviour of users of BCIs will be minimal and observational analysis may not obtain sufficient data to analyze UX. Moreover, severely disabled people, such as patients with locked-in syndrome (LiS) who lose all their muscle control except for vertical eye movements [25] and who constitute a major user group for BCIs, are not able to show any overt behaviour at all. Consequently, in clinical experiments observational analysis is not a strong method for evaluating UX, although for studies in natural environments they might prove useful.

3.2 Neurophysiological Measurement

Task performance metrics have been suggested as quantitative-objective measures of UX but these are not necessarily the indicators of UX. Especially in entertainment applications, there might not be a clear task or users might prefer navigating in the virtual environment without any urge to complete tasks. More recently, use of neurophysiological signals was proposed to model emotional state of users in play technologies [23]. Examples of psychophysiological signals are EEG, galvanic skin response (GSR, measuring skin conductivity) and electrocardiogram (ECG, measuring electrical heart activity). Measured emotions capture usability and playability through metrics relevant to play experience so they provide objective data. They account for user emotion and they are represented continuously over a session. While interacting with a BCI, at least one neurophysiological signal, the EEG, can already be recorded as it is used as an input signal. It is a golden opportunity to extract UX-related features from the brain signals using the same signals. Several problematic issues can be identified when recording psychophysiological signals. Firstly, the sensors attached to the user might induce discomfort to the user, restrict movements or influence the experience. So, the researchers should limit the amount of sensors applied on the user. Secondly, while measuring the UX through the same neurophysiological sensor that is used for controlling the application, UX-related responses should be differentiated from task-related activity.

3.3 Interviewing and Questionnaires

Interviews and questionnaires provide subjective data for assessing UX. They take place after interacting with a system thus are unobtrusive but then not able to extract instantaneous experiences during interaction. For disabled users, especially those with LiS, using subjective methods might not seem to be the easiest way to assess UX as these people might not be able to talk or write. However, if the interviews and questionnaires are prepared in such a way that they can be answered using a few number of choices, such as yes, no and maybe, then they can be completed by these users as well.

Interviewing is a qualitative-subjective technique. During interviews, researchers should be careful to pose the right questions during the interview, if necessary, by monitoring the interaction and detecting unexpected events. The interviewers should remain neutral and refrain from asking leading questions. Questionnaires, on the other hand, are designed to provide quantitative-subjective data. Development of UX questionnaires for entertainment applications has received attention from researchers, especially those who are interested in games. The recently developed Game Engagement Questionnaire [26] includes items related to absorption, flow, presence and immersion. There are also questionnaires focusing exclusively on the components that contribute to UX such as presence [27] and immersion [28]. Another concept that is often related to UX is the usability of the interface. Many heuristics have been proposed for evaluating the usability of video games [29]. However heuristic evaluation does not involve actual users and usability of an interface alone does not represent the UX. Before questionnaires are used

to evaluate BCIs, they may require adaptation taking into account that state-of-the-art BCI applications are relatively simple thus modest in providing rich UX. BCI recognition performance should also be taken into account, as a relatively low performance might influence the UX.

The important factors in selecting the right UX assessment method for BCIs can be listed as the ease of deployment and analysis for the researcher, the comfort of deployment on the user, the strength and reliability in representing the actual UX, and the width of the user spectrum. As seen within this section, all the methods partially fulfill these criteria. Nevertheless, questionnaires stand as strong candidates as they are easy and comfortable to apply, suitable for extracting statistical analyses quickly, strong and reliable when validated and applicable to the majority of the BCI users.

4 Towards a Standardized Questionnaire

As we have described in section 3, classical UX evaluation of a system is often done by administering standardized questionnaires, taking qualitative in-depth interviews with users or by observing overt behavior of users. In this part we will elaborate on our proposal for a standardized questionnaire that can be used to evaluate the UX of BCIs for entertainment purposes. To these means we will first describe some case studies from our group.

4.1 Case Studies

In the past, we have made some prototypes of BCIs incorporated into games that we evaluated on UX.

The first prototype is a BCI game called BrainBasher utilizing the ERD/ERS of imagined and actual movement as described in [14]. Users are provided with direct continuous feedback whether left, right or no movement is detected by the BCI. The game was evaluated with a questionnaire, which compared the UX in the imagined and actual movement conditions. Imagined movement was perceived as more challenging to users although also as mentally more demanding and more tiring. While users preferred the imagined movement for short periods, users would prefer actual movement for prolonged periods of playing this game. Users considered the feedback to be valuable, because they could play around and adapt their strategy of imagining movements to what the system would recognize correctly.

The second study which studied UX and user-centered design involved the game IntuiWoW[30] IntuiWoW is a BCI extension of the popular game World of Warcraft. Users were involved in the design process. They could choose which mental tasks should be used for certain actions in the game. In the subsequent evaluation of the prototype which incorporated those actions users found the recognition accuracy of the mental task to be most important for the user's preferences of mental tasks. In addition, UX was improved with ease of use, fun, intuitiveness and suitability.

Thus, these two studies showed that to evaluate usability of a system it is not sufficient to focus on BCI performance but, that also UX needs to be considered. The first study provides us with a base for developing the questionnaire, the second study provides us with more insight as to what factors are the most important to ask the users.

4.2 Questionnaire

Structured standardized questionnaires can provide valuable quantitative information about an application. One possible cause for the lack of structured questionnaires for BCI applications is that the way of interacting with an application is inherently different from classical human computer interaction. Also it is difficult to compare between different BCIs because of different mental tasks and the way the user uses them. To compare between BCIs and to pinpoint on which dimensions UX can be improved, standardized BCI specific questionnaires are required. Current questionnaires on UX such as the Game Experience Questionnaire [15] and Engagement Questionnaire [26] are not sufficient, because these questionnaires assume that only a traditional method of input (e.g. keyboard and mouse) is used.

As the case studies described in the previous paragraph demonstrated, recognition accuracy, ease of use and applicability of the used mental task play an important role in the UX. Because different mental tasks provide the user with different experiences, it might prove difficult to make every item in the questionnaire relevant. For example, while evaluating how a flickering stimulus for a SSVEP-based BCI is perceived by the user can be very valuable, in the case of an ERD/ERS BCI this is not applicable. However, Zander et al [16] categorize mental tasks used for BCIs into three different groups: passive, active and reactive. Within these groups, mental tasks should, at least for the sake of UX evaluation, be largely comparable. For example, reactive BCIs incorporate stimuli for the user to react to (hence the name 'reactive'). Items in the questionnaire can be on the obtrusiveness of the stimulus, to what extent it distracts attention from the main task of the BCI and to what amount users get tired or sore eyes of it.

Therefore, we propose to develop a questionnaire with modules for each category. Within these categories there can be specific questions about the way the user is interacting with the system. For example, for passive BCIs items in the questionnaire can ask the user if the BCI hardware is comfortable and does not distract from the main task at hand. For active BCIs, items on applicability of the mental tasks and perceived speed of the BCI on the users actions can give valuable information. Also the time that is needed to train the system and the ability to retain that training model over time are important for the UX. To perform an active mental task a user needs a certain amount of concentration. Over time this will fatigue the user and light headaches are not unlikely to develop. When developing and evaluating reactive BCIs, we are more interested in the obtrusiveness of the stimuli that are used. In the case of the aforementioned SSVEP BCI, the flickering of the stimulus is needed to make the BCI work, but

variations in size, colour and texture can make a big difference in how the user perceives the obtrusiveness of the stimulus.

However, there are some difficulties with the categorization of BCIs into 'active', 'passive' and 'reactive'. One can think of BCIs in which the interaction paradigm is ambiguous to two categories.

Consider for example a P300 based lie detector, in which the 'user' gets presented with stimuli and the interogator can find out if the user is interested in what is presented to him, or has seen something before. This is a reactive BCI paradigm used in a passive interaction paradigm, because there is no direct immediate feedback to the user.

Another example is the previously mentioned IntuiWoW [30] in which the user's parietal/occipital alpha power is used to control the form and function of the avatar in the game. Users can actively influence there alpha power by relaxing or agitating themselves, or they can just play the game and let the BCI passively adapt the avatar to their state of mind.

So, to construct our questionnaire, we need to look at the way users interact with the BCI. And while the categorization into passive, reactive and active BCI paradigms provides a good starting point, we need to not just look at how the user performs the mental task, but as well at how the recognition of that mental task is incorporated in the application.

To get BCI research one step further and to bridge the gap between the technology and the user, we need to develop and incorporate standardized measures inspired by HCI. While (neuro) physiological measures are still in (early) development, standardized questionnaires can provide valuable information.

5 Conclusions

In this paper we stressed the need for UX evaluation of BCI applications. While some research has been done on this, it remains largeley an uncultivated area of research. However, we can learn from methods developed in the field of human-computer interaction. Especially standardized questionnaires can be a valuable tool to evaluate BCIs. We proposed a questionnaire in which a core with general questions and modules for the different kinds of mental tasks and ways of interacting with the BCI is developed. This way, we can compare user experience between applications or within one game with different possible input modalities. This can provide developers of BCI applications with valuable information on where the weak and strong spots in the design are, as to create a better UX and acceptance of BCIs.

Acknowledgements

The authors 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. This work was partially supported by the ITEA2 Metaverse1 (www.metaverse1.org) Project.

References

1. Zickler, C., Donna, V.D., Kaiser, V., Al-Khadairy, A., Kleih, S., Kübler, A., Malavasi, M., Mattia, D., Mongardi, S., Neuper, C., Rohm, M., Rupp, R.: BCI applications for people with disabilities: Defining user needs and user requirements. In: Assistive Technology from Adapted Equipment to Inclusive Environments: AAATE 2009, pp. 185–189. IOS Press, Amsterdam (2009)
2. Lightbody, G., Ware, M., McCullagh, P., Mulvenna, M., Thomson, E., Martin, S., Todd, D., Medina, V., Martinez, S.: A user centred approach for developing brain-computer interfaces. In: 2010 4th International Conference on Pervasive Computing Technologies for Healthcare, pp. 1–8. IEEE, Piscataway (2010)
3. Pasqualotto, E., Simonetta, A., Gnisci, V., Federici, S., Belardinelli, M.O.: Toward a usability evaluation of bcis. International Journal of Bioelectromagnetism (to appear 2011)
4. Brooke, J.: SUS: A “quick and dirty” usability scale. In: Usability Evaluation in Industry, pp. 189–194 (1996)
5. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology 52, 139–183 (1988)
6. Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H., Pfurtscheller, G.: Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. IEEE Trans. on Neural Systems and Rehab. Eng. 15, 473–482 (2007)
7. Nijboer, F., Birbaumer, N., Kübler, A.: The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis – a longitudinal study. Frontiers in Neuroscience 4, 55 (2010)
8. Kleih, S., Riccio, A., Mattia, D., Schreuder, M., Tangermann, M., Zickler, C., Neuper, C., Kübler, A.: Motivation affects performance in a P300 brain-computer interface. International Journal of Bioelectromagnetism (2010) (to be published)
9. Rheinberg, F., Vollmeyer, R., Burns, B.: FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. Diagnostica 47, 57–66 (2001)
10. Barbero, A., Grosse-Wentrup, M.: Biased feedback in brain-computer interfaces. Journal of Neuro Engineering and Rehabilitation 7, 34 (2010)
11. Tractinsky, N., Katz, A., Ikar, D.: What is beautiful is usable. Interacting with Computers 13, 127–145 (2000)
12. Moore Jackson, M., Mappus, R.: Applications for brain-computer interfaces. In: Nijholt, A., Tan, D. (eds.) Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction, pp. 89–103. Springer-Verlag London Ltd., London (2010)
13. Oude Bos, D., Reuderink, B.: BrainBasher: a BCI Game. Extended Abstracts of the International Conference on Fun and Games (2008)
14. 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, 33–47 (2010)
15. IJsselstijn, W., de Kort, Y., Poels, K.: The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games (2008) (manuscript in preparation)
16. Zander, T., Kothe, C., Jatzev, S., Gaertner, M.: Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In: Nijholt, A., Tan, D. (eds.) Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction, pp. 181–199. Springer-Verlag London Ltd., London (2010)

17. Solovey, E., Girouard, A., Chauncey, K., Hirshfield, L., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R.: Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In: Proc. of the 22nd ACM Symposium on User Interface Software and Technology, pp. 157–166. ACM, New York (2009)
18. Fairclough, S.: Fundamentals of physiological computing. *Interacting with Computers* 21, 133–145 (2009)
19. Friedman, D., Leeb, R., Guger, C., Steed, A., Pfurtscheller, G., Slater, M.: Navigating virtual reality by thought: What is it like? *Presence: Teleoperators and Virtual Environments* 16, 100–110 (2007)
20. Slater, M., Steed, A.: A virtual presence counter. *Presence: Teleoperators & Virtual Environments* 9, 413–434 (2000)
21. Groenegress, 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, 1–11 (2010)
22. Mandryk, R.L., Atkins, M.S., Inkpen, K.M.: A continuous and objective evaluation of emotional experience with interactive play environments. In: CHI 2006: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1027–1036. ACM, New York (2006)
23. Mandryk, R.L., Inkpen, K.M., Calvert, T.W.: Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology* 25, 141–158 (2006)
24. Fatourechi, M., Bashashati, A., Ward, R.K., Birch, G.E.: EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology* 118, 480–494 (2007)
25. Bauer, G., Gerstenbrand, F., Rumpl, E.: Varieties of the locked-in syndrome. *Journal of Neurology* 221, 77–91 (1979)
26. 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, 624–634 (2009)
27. van Baren, J., IJsselsteijn, W.: Measuring presence: A guide to current measurement approaches, Deliverable 5 for OmniPres project (2004)
28. Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 641–661 (2008)
29. Omar, H., Jaafar, A.: Heuristics evaluation in computer games. In: Bakar, Z.A., Sembok, T.M.T., Zaman, H.B., Bruza, P., Crestani, F., Urs, S.R., Awang, Z. (eds.) 2010 International Conference on Information Retrieval Knowledge Management, pp. 188–193. IEEE, Piscataway (2010)
30. Plass-Oude Bos, D., Poel, M., Nijholt, A.: A study in user-centered design and evaluation of mental tasks for BCI. In: Lee, K.-T., Tsai, W.-H., Liao, H.-Y.M., Chen, T., Hsieh, J.-W., Tseng, C.-C. (eds.) MMM 2011 Part II. LNCS, vol. 6524, pp. 122–134. Springer, Heidelberg (2011)