Chapter 11
Brain–Computer Interfaces and User Experience Evaluation

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

11.1 Introduction

Brain–computer interfaces (BCIs) aim 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 users with and without disabilities. BCIs will be integrated into existing interactive applications. The aim of such applications is to create positive experiences that enrich our lives rather than only providing reliable control. Recently, it was suggested at several keynote presentations at large BCI conferences that reliability is the most important issue to be addressed to achieve technology transfer to the market and the society. However, perfectly reliable systems are not necessarily usable. Even reliable assistive technologies may get abandoned by users when usability is not warranted [35]. Making interactive systems usable is the core expertise of the field of HCI (see Chap. 9). The process of designing interactive systems in the field of HCI consists of analysis of requirements, design and implementation of the system and user evaluation. To evaluate such systems, the user experience (UX) needs a more important role in BCI studies. Researchers should not only focus on the reliability of the control signal, so that we can better understand how such a system can satisfy the needs of the user (see also Chap. 8 for user centered design).

At this point we should make clear that the concept of usability is not the same as user experience, although they are related. The most widely accepted model of measuring user-oriented quality assessment of interactive systems consists of three elements: functionality, usability and the user experience [24]. Functionality...
is about what one can do with the system, id est, what role does it fulfill? Technical aspects such as performance, maintainability, reliability and durability are important. Usability contains higher level concepts such as satisfaction, efficiency, effectiveness, learnability and usefulness. These can partly result from the functionality but are mainly defined by the interaction of the user with the system. Hence, these concepts cannot be tested without real users. User experience is about what the user feels and experiences using the system. User experience therefore contains higher level concepts as immersion (the user is involved and/or lost track of time), fun, engagement, presence (in case of a game, users experience being “in” the virtual world) et cetera. Even though usability and user experience evaluation is not common in current BCI studies, the user’s experience may 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 some case studies and provide recommendations for evaluating user experience with BCIs.

11.2 Current State of User Experience Evaluation of BCI

11.2.1 User Experience Affects BCI

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, develop user requirements, and evaluate the usability [14, 21, 33, 41] and Chap. 8. What is often ignored, however, is the importance of assessing the UX and user acceptance in a structured way during or directly after interaction with the system. The BCI studies that do include UX evaluations indicate three main reasons: its potential to increase user acceptance, to improve performance of the system, and to increase enjoyment. Each of these are discussed in more detail next.

In a study by Münßinger et al., the mood and motivation of users of a BCI painting application was evaluated using a visual analogue scale (VAS) [26]. Patients with amyotrophic lateral sclerosis (ALS) were more motivated to train with the application than healthy users. While the healthy users also had other options for creative expression, this BCI application provided a unique opportunity to the paralyzed patients. Several BCI studies suggest a relation between motivation and BCI task performance [20, 27], and small but significant effects have been found [17] using an adapted version of the Current Motivation Questionnaire. This questionnaire assesses the current motivation in learning and performance situations [34]. Similarly, the users’ belief of how accurately they can control a BCI has an influence on their actual performance. Barbero and Grosse-Wentrup observed 67
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 [1].

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 [38]. This perception could influence motivation which in turn could influence performance. Similarly, a more positive experience may cause users to be more indulgent towards minor usability problems, increasing the user acceptance [29].

Most current BCI applications still serve only as a proof of concept [25], 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 [18,32]. The Game Experience Questionnaire was used to assess immersion, tension, competence, flow, negative affect, positive affect, and challenge [15]. In the first study, the UX and performance were 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.

While more research is still needed, the few studies so far suggest that UX can affect a BCI system in important ways. Therefore it is vital that the UX of BCI systems is properly evaluated.

11.2.2 BCI Affects User Experience

UX can influence the performance of BCIs, but BCIs can affect the UX as well, in two ways: (1) through the effects of using this particular input modality, and (2) 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. Here are some examples to illustrate this.

Using BCI for input can in itself influence the UX (see also Chap. 10). Friedman et al. [9] investigated whether the use of imaginary movement to walk in a virtual world would increase the sense of being present there, using the Slater–Usoh–Steed presence questionnaire combined with a non-structured interview [9,36]. In a follow-up experiment, Groenegress et al. compared the presence experienced with a P300 interface to eye gaze and wand navigation [10]. Both experiments concluded that the BCI did not have a positive influence on presence. In a study by Vilimek and Zander [39], an eye gaze system was augmented with a BCI to simulate the mouse click. The resulting workload of the BCI method was compared to the standard
method of using dwell times for activation, using the NASA TLX [13]. There was no significant difference between the workload for either activation method, so the BCI did not result in a higher cognitive demand. A more recent study by Hakvoort et al. [12] compared a BCI selection method with a non-BCI selection method, made equivalent in terms of time and effort necessary for selection [12]. The comparison was based on affect, evaluated with the self-assessment manikin [4], and on immersion, which was determined with the questionnaire developed by Jennett et al. [16]. In this case, the BCI did turn out to be more immersive and to result in a more positive experience.

With the help of BCI, users can also be supported in the tasks they are trying to accomplish, which in turn should increase user satisfaction. For example, error-related brain activity can be detected and used to fix user or system errors for improved error handling [40]. The amount of information presented on screen can be adjusted according to the user’s workload [37]. BCI could also be used to create or maintain specific user experiences. As an example, brain activity indicators of stress or boredom can be used to keep the user in the optimal state of flow, where the challenge of the task is matched to the skill of the user [6, 7].

But the influence of BCI on UX may extend even further. Obbink et al. [30] investigated the influence of using a BCI on social interaction in a cooperative game [30]. The social interaction was assessed in terms of the amount of speech, number of utterances, and gestures. Additionally, a custom questionnaire at the end of the experiment was provided to evaluate the participants’ self-reported, subjective experience. Because of the higher difficulty of the BCI-based selection, compared to point-and-click with a traditional mouse, there were more utterances and empathic gestures (see also Sect. 11.4).

All in all, whether BCI is used to affect the UX purposefully or whether this happens by simply using this input modality, in both cases it is important to evaluate and be aware of the effects. The next section will show different methods to do this, and discuss the implications for using them to evaluate BCIs specifically.

11.3 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. 11.1). 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...
Fig. 11.1 A classification of current user experience evaluation methods used in human–computer interaction for entertainment technologies (adapted from [22])

the quadrants formed by these two axes and discuss their contribution in evaluating BCI systems.

11.3.1 Observational Analysis

Observational analysis is a qualitative–objective method which relies on overt user response. 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 a noise-free environment during audio recording or consistent illumination during video capturing. Some restrictions are also imposed by brain activity recording devices. For example, the electroencephalogram (EEG, measuring electrical brain activity) is affected by the user’s movement [8], so users are usually asked to keep their bodies and faces 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 [3] and who constitute a non-negligible 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 it might prove useful.

11.3.2 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 the 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. First of all, the research on using neurophysiological sensors to measure UX is in its infancy. The neurophysiological correlates of UX or its components are not well-defined which makes this method rather a questionable one. Secondly, the sensors attached to the user might induce discomfort to the user, restrict movements or influence the experience. So, the researchers should limit the number of sensors applied on the user. Thirdly, 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.

11.3.3 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. One way to converge capturing short-term UX might be to conduct questionnaires and interviews incrementally, instead, in multiple sessions, rather than conducting a single questionnaire/interview after the interaction has taken place. 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 small 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. An example demonstrating the use of interviews in BCI UX evaluation is the study by Gürikök et al. [11]. In their study, the authors conducted interviews with participants to find out the reasons why people switched between BCI and speech control in a multimodal game.

Questionnaires are designed to provide quantitative–subjective data. Users rate the items in a questionnaire on a Likert-scale or a Visual Analogue scale, which yields a number of how much they agreed with a statement. 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 [5] 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 [2] and immersion [16].

11.3.4 Other Methods

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 [31]. However heuristic evaluation does not involve actual users and the 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.

Analysing logged software data is also considered as a quantitative–objective method for UX evaluation in some studies. Logs are not direct correlates of UX but they might be helpful in understanding the course of interaction, identifying problems or certain preferences, and thus in designing for better UX. For example, by analysing the frequency of key presses in a game, one can derive a cluster of events to which the player was more reactive and can use this new information to design better interaction.

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.
In this section we will elaborate on two case studies in which we applied various methods of UX evaluation and will try to explain why we chose a certain method and how it answers our research questions (see other experiences in Chap. 8).

### 11.4.1 Case Study: Mind the Sheep!

We did a series of UX evaluation studies using the multimodal game we developed, called Mind the Sheep!. The game world (see Fig. 11.2) is a meadow on which a number of (white) sheep move autonomously and the (black) shepherd dogs can be commanded by the player. When a dog approaches some sheep, the sheep will tend to flock and move away from the dog. This way the sheep are herded in a desired direction. The goal of the game is to gather the sheep in a pen as quickly as possible.

The game can be played using different modalities in different ways. In the BCI controlled version of the 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 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 EEG 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.

In the first study we describe here, we compared BCI control to simple mouse control to study the social interaction between players in the cooperative multimodal version of the game [30]. To control the dogs using the mouse, the player first clicks on the dog they want to select and then on the location they want the dog to move to. In the cooperative multiplayer version of the game, co-located players work together to pen the sheep so they need to interact while playing to develop a strategy. However, interaction means such as speech and bodily movements might impair the accuracy of the BCI due to the noise they impose on the EEG. So there is a trade-off between maintaining a strategy during the game and maintaining a certain accuracy level. We did an experiment with ten pairs playing the game with both controllers. We performed an observational analysis of the audio–visual data recorded during the play. Though non-significantly, during mouse control the participants produced more speech and instrumental gestures, which are overt messaging channels. This implies that they interacted more freely with mouse control. On the other hand, during BCI control participants produced, again non-significantly, more utterances and empathic gestures, which are emotion signalling channels. This finding suggests that the participants were affected more by the events during the BCI game; perhaps they were surprised, or things went wrong more often.

In another study with the single-player version of Mind the Sheep!, we evaluated UX in terms of immersion and affect through questionnaires [12]. Again, we compared BCI control to mouse control but this time the way the mouse was used was different. Now, the player had to hold the mouse button pressed when they wanted to make a selection. The dogs were highlighted one at a time with an increasing highlight period. When the player released the mouse button, the currently highlighted dog was selected. To make an accurate selection, the player needs to react in the time when the dog they want to select is highlighted. This way, mouse control becomes similar to BCI control so that they both offer some challenge to the player. In our experiment we let 17 participants play the game with BCI and mouse control and after each game we evaluated UX using the self assessment manikin [4] for affect and the immersion questionnaire [16]. Evaluation results showed that BCI control was found to be more immersive ($p = 0.031$) and positively affective ($p = 0.044$) than mouse control. Furthermore, analysis of the logged game data revealed that participants appeared to have more patience with BCI control than mouse control, which could have been caused by the curiosity of participants for BCI control or by their self overestimation during mouse control.
11.4.2 Case Study: Hamster Lab

The aim of the study was to investigate the effect of level of control on user experience. We conducted an experiment, with 200 participants, in which they played a game with a varying amount of control: the game is called Hamster Lab [19]. The user controls a hamster situated in maze-like game levels, set in a laboratory setting (see Fig. 11.3). The user controls the hamster by pressing the arrow keys on a keyboard. This way, the user has five possible options for control: up, right, down, left and do not move. The 15 controlled conditions in the game are specified with a certain amount of control. From perfect control, where every press of the button is directly translated to the corresponding action in the game, to 20% control, that is, chance level, where there is an even chance for a certain action to be translated to any of the five possible actions. Using this manipulation of control we can simulate the feeling of unreliable input such as is the case with a BCI. Thus, whereas the relation between control and user experience can only be investigated through correlational analysis in conventional BCI experiments, through this simulation we can study the effects using controlled conditions. Hamster Lab is an online game suitable for playing in a web browser. This made it easier to gather participants for the study we will next describe.

This study was conducted to find the relationship between fun and control. After the user commanded the hamster to the exits of the four mazes a short questionnaire was presented to the user. The questionnaire was kept to a bare minimum of what we wanted to know. There were nine questions in total: six were Visual Analogue Scale (VAS) items and three questioned for basic demographics (age, gender and a field where user could give feedback. As it was an online game probably not all participants were motivated to answer the number of questions one would normally ask after an experiment. Furthermore, we wanted the participants to play multiple sessions to gather more data. Based on IP addresses 200 unique participants started a round. Three hundred and fifty-one rounds in total were started. Two hundred and twelve (60.4%) of these runs were continued through the four levels and filled in the questionnaire completely. By most participants the short questionnaire was appreciated (12 people played five rounds of four levels). Though for 39.6% of the runs, even nine questions were too much for the participant to bother, as some entries in our database showed with comments like “Why these questions, I want to play!” and items that remained unanswered. This number includes the runs that were started and terminated halfway, for example, participants that did not like the game and closed their browser window.

The six VAS items provided us with information on the UX of the participants on the following concepts: fun, frustration, control, dominance and empowerment. First we had to assert that our way of influencing the amount of control was also perceived by the user as such. Regression analysis on the amount of control and the perceived control showed a very significant linear trend, hence, the higher the amount of control “given” to the user, the higher the amount of control they
perceived. For the relationship between fun and control we hypothesized that fun is positively influenced by control (users are able to do the things they intend) but an optimum exists before the maximum amount of control (the game is too easy and/or nothing surprising happens). A regression analysis showed that more variance was explained using a third order polynomial (34.9%) than using a linear (29.1%) trend would. The third order polynomial showed an optimum before the 100% control mark which was also shown by comparing the medians of the conditions with a high amount of control. This supports our hypothesis and is also in line with the idea that users playing a game want to have some kind of challenge in a game. In this way, the unreliable input from a BCI can be used as a challenge in a game [28].

The result is validated only for this specific game, other games might show a curve looking slightly different but with the same characteristics. Games that are in itself a big challenge for the user might require 100% control. An interesting conclusion from this is that one wants to use a BCI for control, the game difficulty needs to be adjusted to balance the user skills and game challenge for the optimal experience of flow [6].
11.5 Discussion and Conclusion

In this paper we stressed the need for UX evaluation of BCI applications. While some research has been done on this, it remains largely an uncultivated area of research. However, we can learn from methods developed in the field of human–computer interaction.

To evaluate a BCI system several methods are available: observational analysis can be used in settings where the interaction of the user(s) with the BCI system as a whole is important. For example in the case study of Mind the Sheep! we showed that observational analysis is a useful method when evaluating systems in a realistic setting especially when users can also interact with each other. When overt user response to the system is limited, id est, in case of a clinical experiment, when the user is disabled, or in non observable settings such as for example a web-based experiment, observational analysis is less useful.

Neurophysiological measurements are a quantitative–objective method to assess UX. However, these techniques are still topic of research and most are not very reliable at the present. If a reliable neurophysiological method is used however, this provides a worthwhile source of information as the signal is continuous of nature, as opposed to for example a questionnaire.

Interviews are especially useful in explorative studies. Asking open (non-leading) questions can lead to the reason why a user does or does not like a certain aspect of the system or why the users did what they did. This information is hard to capture through other methods, as it is quite detailed in nature.

Questionnaires are quantitative of nature and answers to the questionnaire can easily be quantified to prove effects over groups of participants. This makes it a frequently used method to evaluate systems. Standardized questionnaires exist on various aspects of UX. However, if one wants to evaluate a system on all these aspects the user has to fill in hundreds of questions, with the risk of “questionnaire fatigue” (filling in the questionnaire at random, or the same answer for each item) and users choosing for the safe middle option, because at some point, all questions seem to be the same. In the case study Hamster Lab, we showed that when a high number of participants is involved or data is gathered over multiple sessions it is possible to limit the items in the questionnaire to exactly what is needed to answer your research question.

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 Project (http://www.metaverse1.org).
References

AUTHOR QUERIES

AQ1. First author has been considered as corresponding author. Please suggest.
AQ2. Please provide complete details for ref. [12, 15]
AQ3. Please provide volume number for ref. [14], if available
AQ4. Please provide page range for ref. [26]
AQ5. Please provide publishing year and publisher location details for refs. [30, 33]
AQ6. Please provide publisher location for ref. [35]