User expectations and experiences of a speech and thought controlled computer game

Hayrettin Gürkök, Gido Hakvoort, Mannes Poel, Anton Nijholt
Human Media Interaction
University of Twente
Enschede, the Netherlands
{h.gurkok,m.poel,a.nijholt}@utwente.nl, gido.hakvoort@gmail.com

ABSTRACT

Brain-computer interfaces (BCIs) are often evaluated in terms of performance and seldom for usability. However in some application domains, such as entertainment computing, user experience evaluation is vital. User experience evaluation in BCI systems, especially in entertainment applications such as games, can be biased due to the novelty of the interface. However, as the novelty will eventually vanish, what matters is the user experience related to the unique features offered by BCI. Therefore it is a viable approach to compare BCI to other novel modalities, such as a speech or motion recogniser, rather than the traditional mouse and keyboard. In the study which we present in this paper, our participants played a computer game with a BCI and an automatic speech recogniser (ASR) and they rated their expectations and experiences for both modalities. Our analysis on subjective ratings revealed that both ASR and BCI were successful in satisfying participants' expectations in general. Participants found speech control easier to learn than BCI control. They indicated that BCI control induced more fatigue than they expected.

Categories and Subject Descriptors

B.4.2 [**Input/Output Devices**]: Channels and controllers; H.5.2 [**User Interfaces**]: Input devices and strategies

General Terms

Experimentation, Human Factors

Keywords

Brain-computer interface, automatic speech recogniser, user experience, expectations

1. INTRODUCTION

Brain-computer interfaces (BCIs) can translate brain signals directly into computer commands in order to provide a muscle-independent control. BCIs have been proven to be successful input modalities for various computer applications such as games and virtual reality applications [15]. Although they are not as perfect as the keyboard or mouse in terms of reliability, they are valuable controllers in entertainment computing because gamers enjoy interacting through novel technologies and tackling the challenges caused by the shortcomings of these technologies [12]. Lately we are witnessing a high interest in BCI controlled games from researchers as well as from the public, especially with the emergence of affordable BCI hardware. The most popular genre of such games are the emotion-attention based games played by using portable hardware such as the NeuroSky MindSet[1] or the Emotiv EPOC[2].

A traditional approach to evaluate BCIs is to assess their performance, for example in terms of recognition accuracy or mutual information [14]. However the performance of a system does not necessarily imply user satisfaction. The system might recognise user actions perfectly but might be difficult to use. For this purpose, some methods to evaluate the ease of use (i.e. usability) of BCIs [11, 13] have been proposed. However, usability alone does not imply user satisfaction either. Especially in entertainment technologies factors such as fun, affect, engagement and immersion play a crucial role in user satisfaction. Some of these factors have also been evaluated before. For example BCI based control of a game was found to be more immersive and positively affective than the mouse based control of the same game [4]. A question that arises is whether such a finding is due to the novelty of the BCI compared to the mouse or due to the unique possibilities offered by the BCI. A viable approach to answering this question is to compare BCI to a non-traditional modality such as a speech or motion recogniser.

An automatic speech recogniser (ASR) [6] is a non-traditional input modality, as is a BCI, and is often used for hands-free control in cars, mobile phones, and some day-to-day applications. Using an ASR is intuitive in the sense that the action required to use this interface, speaking, is what we mainly use when we interact with each other in the real world. ASR is a good alternative to BCI for investigating user experience in games because it is still a novel game controller for many people.

In this paper we evaluate user experience in a computer game, in relation to user expectations, to see whether a steady-state visually evoked potential (SSVEP) based BCI is suitable as a game controller. We base our user expe-

[1]http://store.neurosky.com/collections/games
[2]http://www.emotiv.com/store/apps/applications/117/

Full presentation, ACE'2011 - Lisbon, Portugal

rience evaluation on subjective ratings of the factors speed, pleasantness, accuracy, fatigue, learnability, naturalness and enjoyability. We also evaluate ASR and compare it to the SSVEP based BCI to explore on which aspects the two novel modalities differ in terms of user experience. As experimental platform we use a multimodal computer game controlled by making selections as selection is a typical task for both modalities, for example in smart homes [2, 10].

The rest of this paper is organised as follows. In section 2, we introduce BCIs based on the SSVEP and report research related to our study. Then, in section 3, we describe our method and tools for this study. Section 4 describes our experiment details and analysis results. In section 5 we discuss these results and in section 6 we conclude by re-stressing the major findings.

2. BACKGROUND

2.1 BCIs Based on SSVEP

BCIs can infer a user's intention by interpreting brain activity. 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.

SSVEP is a stimulus dependent, widely used neuromechanism. When a person attends to a visual stimulus that is repeating with a certain frequency, the amplitude of the signal measured from the visual cortex is enhanced at the frequency of the stimulation. This enhancement is known as the SSVEP [7]. 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 flicker frequency, size, colour and shape of the stimulus [1].

2.2 User Experience Evaluation in Multimodal Systems

As we already mentioned in section 1, some studies previously focused on evaluating BCI and ASR systems in terms of usability through questionnaires. A good number of questionnaires were proposed and validated to evaluate usability although there is no consensus among the standard usability evaluation questionnaires for multimodal systems [18]. Moreover, usability is not the sole indicator of user experience. In human-computer interaction research, usability and user experience are two concepts which are defined and evaluated separately. Usability is defined as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*" [8]. User experience is still related to the pragmatic quality of a system, just as usability is, but it is also concerned with the hedonic quality of a system. The hedonic quality of a system is about providing stimulation, communicating identity, and provoking valued memories [5].

User experience is influenced by users' values, abilities, prior experiences and knowledge as well as the context of use. Every experience a user has with a product affects not only their next experience with the product but also their future experience with any product using the same technology as control input. So, a product can change the conceptions or conclusions about a technology. If the change is in a positive way, then we can say that the product provides a positive user experience. To give an example, let us consider the BCI technology which has long been used in assistive systems. For a non-disabled person, controlling a wheelchair by imagining hand movements might seem to be difficult and inefficient so they might have low expectations on these aspects of BCIs. But when they play a motor imagery based car racing game they might consider the difficulty of BCI as a challenge they enjoy tackling. The key to success is to find the right interaction design which enables the technology to improve upon user's expectations. Therefore, it is a viable approach to assess user experience with respect to user expectations (thus previous experiences), especially when comparing multiple systems.

There are not many user experience evaluation methods suitable for multimodal entertainment systems. One prominent method is the SUXES [16] which measures user experience with respect to user expectations. Its use in evaluating assistive multimodal systems [17] as well as BCI systems [3] was demonstrated in previous studies. In this study we will use this method with some modifications on the questionnaire it contains. We will describe the modified method in detail in section 3.2.

3. TOOLS AND METHOD

3.1 The Game

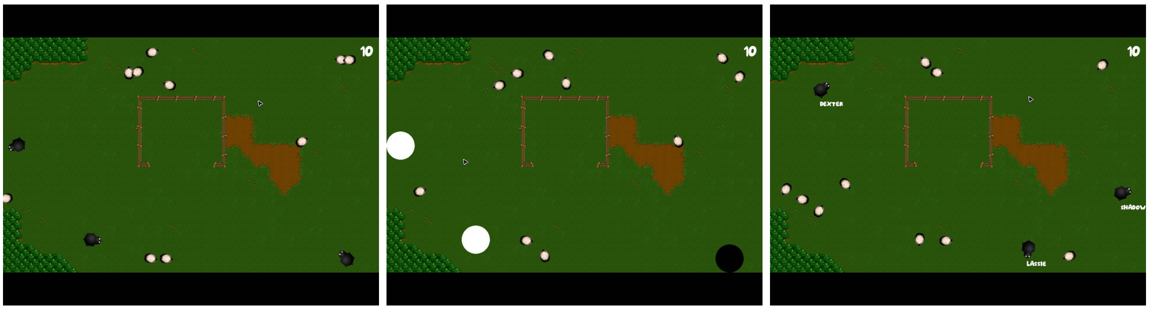
The game that we evaluated in this study is called Mind the Sheep! (see Fig. 1). This is a multimodal computer game where the player needs to herd a flock of sheep across a field by commanding a group of dogs. The game world contains three dogs, ten sheep, a pen and some obstacles. The aim is to pen all the sheep as quickly as possible. For the purpose of this work, we used the BCI and ASR controlled versions of the game. The ASR and BCI are used to select a dog while the directions to the selected dog are given with the mouse.

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 (also see Table 1).

Let us stress that it is the player's mouse press which de-

Table 1: Table explaining user interface response to user actions and the game flow

Triggering user action	No action	Click on the location they want to send a dog and hold mouse button pressed	Release the mouse button
Subsequent user interface response	Dogs are stationary, sheep graze	*ASR game:* No change, *BCI game:* Dog images are replaced with flickering circles	Selected dog moves to the location of the cursor at the time of mouse release
Subsequent user action	None	*ASR game:* Pronounce the name of the dog they want to select, *BCI game:* Concentrate on the circle replacing the dog they want to select	None



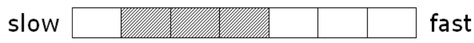
(a) (b) (c)

Figure 1: Screenshots from the game. In (a) BCI game with SSVEP stimulation off. Black images are the dogs and white images are the sheep. In (b) BCI game with SSVEP stimulation on. Dog images are replaced by flickering circles. In (c) ASR game. Names are under the dog images.

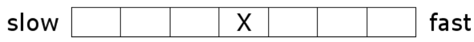
termines when to start and end the data acquisition. In other words, the amount of accumulated data depends on the player. In the BCI game, the longer the user waits, the higher is the accuracy of signal classification. But meanwhile, the positions of the sheep change so the player needs to trade off between speed and accuracy. In the ASR game this is less of an issue because the player would release the mouse as soon as they have pronounced the dog's name once.

3.2 User Experience Evaluation

For user experience evaluation, we used the method proposed in [3]. Although the use of this method was demonstrated for a BCI game, it is based on the generic SUXES method [16] which can be used to evaluate multimodal systems in general. The method works as follows. Before using the system, the user fills in an expectations questionnaire. Each item in the questionnaire is rated through a 7-point semantic differential scale which is anchored by opposite phrase pairs at the ends (see Figure 2(a)). The participants can then indicate their zone of expectations (ZoE) for each item by shading the box scale. Other than the phrase pairs, the scale contains no additional anchoring. It is expected that if in the experiences questionnaire the user marks their experience for an item lower than the ZoE they were disappointed and if they mark it higher they were (positively) surprised. The expectations questionnaire contains the following items and corresponding phrase pairs in parentheses (in the given order): speed (slow–fast), pleasantness (pleasant–unpleasant), accuracy (erroneous–error-free), fatigue (tiring–effortless), learnability (easy to learn–hard to learn), naturalness (natural–unnatural) and enjoyability (boring–fun). In accordance with the discussion we had in section 2.2 the items speed, accuracy, fatigue and learnability capture the pragmatic quality of the system while the items pleasantness, naturalness and enjoyability capture the hedonic quality.



(a) Expectations with respect to speed



(b) Experience with respect to speed

Figure 2: Interpreting expectations and experiences: (a) implies that the user will be surprised if the interface is faster than level 4 and will be disappointed if it is slower than level 2. So the zone of expectations (ZoE) is <2,4> (b) indicates that the user rated the speed as level 4 which is within the ZoE thus meets the expectations.

After using the interface, the user fills in the experiences questionnaire. The questionnaire is identical to the expectations questionnaire but in this case the user does not shade the boxes but puts a cross inside the box that represents their experience (see Figure 2(b)).

4. EXPERIMENT

4.1 Participants

Fourteen people (2 female) participated in the experiment. They had an average age of 24.5 ($\sigma = 2.88$), ranging from 19 to 28 years. None of them were native English speakers. Four of them had previous experience with BCIs and nine of them with ASRs. Four 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 Balancing ASR and BCI Recognition Performances

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.

Seven non-native English speakers participated in the first pilot study in which we collected data to compare ASR performance among different sets of dog names (trios formed by candidate names Hector, Victor, Dexter, Pluto, Shadow and Lassie). To record speech, we placed the microphone behind the participant because when the microphone was in the front, ASR performance was so high that it could not be matched by the BCI. The participants pronouced each name five times. For each name trio we computed the average recall of the recognition carried out by the ASR.

In the second pilot study another 7 people participated. This time we collected data to evaluate BCI performance with respect to different sets of frequencies (trios formed by candidate frequencies 6 Hz, 6.67 Hz, 7.5 Hz, 8.57 Hz, 10 Hz, 12 Hz, 15 Hz). The setup and procedure for this pilot study is described elsewhere in detail [4]. Just as in the first pilot study, for each frequency trio we computed the average recall of the recognition performed by the BCI.

We sorted the name trio-frequency trio pairs with respect to their similarity in average recall, which was assessed by Wilcoxon rank-sum tests (higher p-values indicated more similarity). Among the most similar name trio-frequency trio pairs, the pair with the highest average recall was selected. In this way, we decided to use Dexter, Lassie and Shadow (yielding an average recall of 83%) as dog names and 7.5 Hz, 10 Hz and 12 Hz (yielding an average recall of 84%) as flicker frequencies. This pair yielded a p-value of 0.97. We set the flicker diameter length to 3 cm. 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. The order of the games were counterbalanced among the participants to diminish the effect of familiarity with the game. Before each game, based on their current knowledge and previous experiences, they filled in the expectations questionnaire to indicate their ZoE for selecting dogs using BCI or ASR. They were instructed to shade any number of boxes (between 1 and 7) they wished to, with respect to the devices they would need to use and tasks they would need to do to select a dog. After that, the experimenter collected the questionnaire, left the room and the game began. The participant played each game until all the sheep were penned or the play time reached 10 minutes. After the game, they filled in the experiences questionnaire.

Sound was acquired by the microphone located to the right, behind 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 the participant's head at locations PO3, O1, Oz, O2 and PO4 according to the international 10-20 system [9]. 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

Our analyses were based on the expectations and experiences questionnaires corresponding to BCI and ASR control in Mind the Sheep!. For each item, we computed the medians of the experienced levels and the lowest and the highest expected levels across the participants. We also computed the medians across all the items as the indicator of overall user experience. We compared the difference in expectations and experiences for the two modalities. The significance of differences were assessed by the Wilcoxon signed-rank test ($p < 0.05$).

For each item and for the overall user experience we computed two measures: measure of adequecy (MoA) and measure of superiority (MoS). MoA is the difference between the experienced level and the lower end of the ZoE while MoS is the difference between the experienced level and the higher end of the ZoE. If the experience is within the expectations, then the MoA is non-negative and MoS is non-positive. If the experience is below the expectations then the MoA is negative. If the experience surpassess the expectations then the MoS is positive. In the example in Figure 2, the ZoE is $<2,4>$ and the experienced level is 4. So MoA is 2 and MoS is 0 indicating that the experience was within the expectations. Based on MoA and MoS measures we draw conclusions on the suitability of ASR and BCI as direct control modalities in our game.

4.5 Results

Figure 3 displays the median ZoE and experience values across all participants for BCI and ASR control. As can easily be seen in this figure, the expectations and experiences differ between the two modalities. However the significance analysis of the underlying data reveals that the differences are not significant, except for the experience values for learnability.

The MoA and MoS values computed as described in section 4.4 are given in Table 2. Note that the only non-positive MoA value (i.e. experience falling below expectation) is for fatigue with BCI and the only non-negative MoS values (i.e. experience exceeding expectation) are for the learnability and naturalness with ASR.

5. DISCUSSION

The results of our experiment showed that the expectations (grey cells in Figure 3) of participants from BCI and

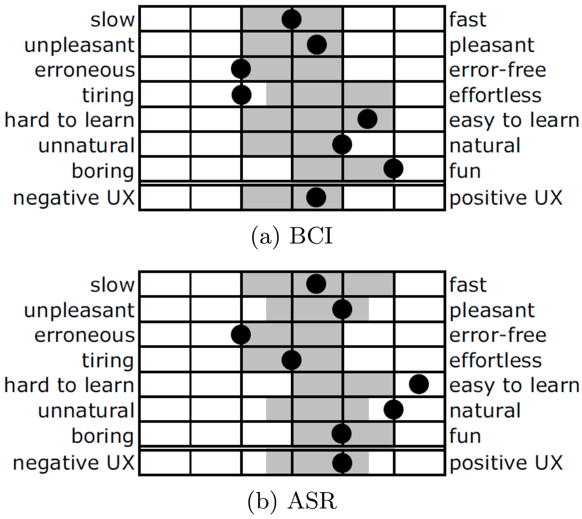


Figure 3: Median values across all participants for ZoEs (grey cells) and experiences (black circles).

Table 2: Measures of adequacy and superiority, MoA and MoS respectively, for BCI and ASR

	BCI		ASR	
	MoA	MoS	MoA	MoS
Speed	1.0	-1.0	1.5	-1.5
Pleasantness	1.5	-0.5	1.5	-0.5
Accuracy	0.0	-2.0	0.0	-2.0
Fatigue	-0.5	-3.0	1.0	-1.0
Learnability	2.5	-0.5	2.5	0.5
Naturalness	2.0	0.0	2.5	0.5
Enjoyability	2.0	0.0	1.0	-1.0
Overall user experience	1.5	-0.5	1.5	-0.5

ASR did not differ significantly per item or on average over all items. Also their experiences for both modalities (black circles in Figure 3) were non-significantly different except that they found ASR significantly easier to learn than BCI. This can be explained by the fact that they found speech input, though non-significantly, more natural to use than BCI input consequently requiring less training or adaptation. We see that learnability and naturalness were rated higher for ASR than for BCI. This is not surprising considering the greater familiarity we have with speaking than we have with concentrating on flickering images. There are various possible reasons why BCI was rated higher for enjoyability. For example it might be due to the novelty of the BCI or the challenge it provided. A deeper analysis is necessary to draw a definitive conclusion.

If we consider the median of the rating scale (i.e. level 4) as a neutral experience, then by looking at the experiences only, we can say that for both modalities the accuracy and for BCI the fatigue were the negatively experienced items. Although we picked the optimal parameters for the accuracy of the BCI, it was not error-free as it is almost impossible to achieve a BCI functioning perfectly for everybody. The equal user experience rating for the accuracy in ASR and BCI supports the validity of the parameter tuning we performed to equalise the performances of the two modalities. The negative experience for fatigue in BCI can be due to the SSVEP stimulation (i.e. flickering circles) during the BCI game, given that the accuracy and speed of both modalities were rated alike. In terms of pleasantness, learnability, naturalness and fun both modalities yielded good user experience. Note that all items pertaining hedonic quality were rated positively for both modalities while for some pragmatic quality items the ratings were negatively. This shows that the game was able to affect the participants positively despite its imperfect usability.

When we analyse experiences with respect to expectations we see that the participants found BCI control more tiring than they expected which can be due to the SSVEP stimulation, as we discussed above. On the other hand, they found speech input easier to learn and more natural than they expected. When we look at the MoA and MoS values of overall user experience for both modalities, we see that they are equal, meaning that both modalities provided a satisfactory overall user experience.

A limitation of our user experience evaluation method is that in some cases it might become overly optimistic. Let us assume the case that the users are evaluating a system for its accuracy and that the technology behind the system is intrinsically an erroneous one. Especially if the users have expertise with this technology, they would rate their expectations for the accuracy of the system low. The typical system based on this technology would function matching their expectations so users would rate their experience for the accuracy also low. At this point, our evaluation method would conclude that the system is accurate enough to satisfy users. This is not a wrong conclusion but an incomplete one. This conclusion explains where this particular system is located among its competitors using the same technology. However it hides the information on how accurate the users found the system, independent of their expectations. This might mislead the evaluators such that they would think that there is no room, at least no need, for improving the accuracy of this system. Especially for a commercial system produced for users with a broad range of expertise levels, this might result in a serious failure on the market. To prevent such cases, the evaluation should be done with user groups balanced with respect to their expertises.

6. CONCLUSION

In this paper we explored whether BCI and/or ASR are suitable modalities for direct control in computer games. We built a multimodal computer game in which the players can make selections using an SSVEP based BCI or an ASR. In an experiment we let the participants play the game using each modality and indicate their pre-game expectations and post-game experiences using questionnaires. We compared their questionnaire answers for each modality. Then we assessed the suitability of each modality by studying the gap between participants' experiences and expectations.

The experiment results showed that the expectations and experiences of participants did not differ significantly, except that they found ASR easier to learn than BCI. This might be an implication of speech input being rated more natural than BCI in providing commands. The hedonic quality of both modalities were rated positively while for some pragmatic quality items the ratings were negative. When experiences are evaluated with respect to expectations, BCI induced more fatigue than expected. This might be explained by the SSVEP stimulation during the BCI game. Speech input was found easier to learn and more natural than expected. It was also rated higher than BCI on these aspects. As people have more experience in speaking than in concentrating on flickering images, it is reasonable that they found ASR control more natural and easier to learn than BCI. Perhaps another type of BCI (i.e. other than an SSVEP based BCI) could have been rated higher. Nevertheless, participants' overall experiences were within their expectations for both modalities meaning that both ASR and BCI were found satisfactory and suitable for direct control in our game.

Our future research direction would be to build BCI games considering our findings and evaluate them following the method we proposed in this paper. We think that an SSVEP-based BCI should have a more intuitive function rather than making selections, for example monitoring concentration level of a player. Other sensory channels should also be considered to evoke steady-state potentials. Players might find audio or tactile stimulation less tiring and easier to use than the visual one but this requires further investigation.

7. 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.

8. REFERENCES

[1] J. Bieger and G. G. Molina. Light stimulation properties to influence brain activity: A brain-computer interface application. Technical Report TN-2010-00315, Philips Research, Eindhoven, the Netherlands, 2010.

[2] G. Edlinger, C. Holzner, C. Guger, C. Groenegress, and M. Slater. Brain-computer interfaces for goal orientated control of a virtual smart home environment. In *4th International IEEE/EMBS Conference on Neural Engineering*, pages 463–465. IEEE, Piscataway, NJ, USA, 2009.

[3] H. Gürkök, G. Hakvoort, and M. Poel. Evaluating user experience with respect to user expectations in brain-computer interface games. In *Proceedings of the 5th International BCI Conference*. TU Graz, Graz, Austria, 2011. To appear.

[4] G. Hakvoort, H. Gürkök, D. Plass-Oude Bos, M. Obbink, and M. Poel. Measuring immersion and affect in a brain-computer interface game. In *Human-Computer Interaction - INTERACT 2011, 13th IFIP TC 13 International Conference*. Springer, 2011. To appear.

[5] M. Hassenzahl. The thing and I: understanding the relationship between user and product. In *Funology: from usability to enjoyment*, pages 31–42. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2004.

[6] J.-P. Haton. Automatic speech recognition: A review. In *Enterprise Information Systems V*, pages 6–11. Springer, Dordrecht, the Netherlands, 2005.

[7] C. S. Herrmann. 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.

[8] ISO 9241-11:1998. *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. ISO, Geneva, Switzerland, 1998.

[9] H. H. Jasper. Report of the committee on methods of clinical examination in electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology*, 10(2):370–375, 1958.

[10] I. McLoughlin and H. R. Sharifzadeh. Speech recognition for smart homes. In *Speech Recognition, Technologies and Applications*, pages 477–494. I-Tech, Vienna, Austria, 2008.

[11] C. S. Nam, Y. Li, Y. Jeon, Y.-J. Kim, and H.-Y. Yoon. Usability of the P300 speller: Towards a more sustainable brain-computer interface. *e-Minds: International Journal on Human-Computer Interaction*, 1(5), 2009.

[12] A. Nijholt, B. Reuderink, and D. Oude Bos. Turning shortcomings into challenges: Brain-computer interfaces for games. In *Intelligent Technologies for Interactive Entertainment*, pages 153–168. Springer-Verlag, Berlin/Heidelberg, Germany, 2009.

[13] E. Pasqualotto, A. Simonetta, V. Gnisci, S. Federici, and M. O. Belardinelli. Toward a usability evaluation of BCIs. *International Journal of Bioelectromagnetism*, 13(1), 2011. To appear.

[14] A. Schlögl, J. Kronegg, J. E. Huggins, and S. G. Mason. Evaluation criteria for BCI research. In *Toward Brain-Computer Interfacing*, pages 327–342. The MIT Press, Cambridge, MA, USA, 2007.

[15] D. Tan and A. Nijholt, editors. *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. Springer-Verlag, London, UK, 2010.

[16] M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and J. Hella. SUXES - user experience evaluation method for spoken and multimodal interaction. In *10th Annual Conference of the International Speech Communication Association*, pages 2567–2570. ISCA, 2009.

[17] M. Turunen, H. Soronen, S. Pakarinen, J. Hella, T. Laivo, J. Hakulinen, A. Melto, J.-P. Rajaniemi, E. Mäkinen, T. Heimonen, J. Rantala, P. Valkama, T. Miettinen, and R. Raisamo. Accessible multimodal media center application for blind and partially sighted people. *Computers in Entertainment*, 8:16, 2010.

[18] I. Wechsung and A. Naumann. Evaluation methods for multimodal systems: A comparison of standardized usability questionnaires. In *Perception in Multimodal Dialogue Systems*, pages 276–284. Springer, Berlin/Heidelberg, Germany, 2008.