How Much Control Is Enough? Influence of Unreliable Input on User Experience

Bram van de Laar, Danny Plass-Oude Bos, Boris Reuderink, Mannes Poel, and Anton Nijholt

***Abstract*—Brain–computer interfaces (BCI) provide a valuable new input modality within human–computer interaction systems. However, like other body-based inputs such as gesture or gaze based systems, the system recognition of input commands is still far from perfect. This raises important questions, such as what level of control should such an interface be able to provide. What is the relationship between actual and perceived control? And in the case of applications for entertainment in which fun is an important part of user experience, should we even aim for the highest level of control, or is the optimum elsewhere? In this paper, we evaluate whether we can modulate the amount of control and if a game can be fun with less than perfect control. In the experiment users (*n* = 158) played a simple game in which a hamster has to be guided to the exit of a maze. The amount of control the user has over the hamster is varied. The variation of control through confusion matrices makes it possible to simulate the experience of using a BCI, while using the traditional keyboard for input. After each session the user completed a short questionnaire on user experience and perceived control. Analysis of the data showed that the perceived control of the user could largely be explained by the amount of control in the respective session. As expected, user frustration decreases with increasing control. Moreover, the results indicate that the relation between fun and control is not linear. Although at lower levels of control fun does increase with improved control, the level of fun drops just before perfect control is reached (with an optimum around 96%). This poses new insights for developers of games who want to incorporate some form of BCI or other modality with unreliable input in their game: for creating a fun game, unreliable input can be used to create a challenge for the user.**

***Index Terms*—Brain-computer interfaces, computer interfaces, human computer interaction.**

I. Introduction

RECENT DEVELOPMENTS in interfaces show that there is a need for less artificial means of control. The most prominent examples of the moment are the Nintendo Wii and the Microsoft Kinect, both gesture interfaces. However, speech, eye gaze, and other physiological measures are also promising a more intuitive way of interaction. By allowing the user to apply knowledge from previous interactions, for example from interacting with the real world or from interacting with comparable systems, the interface is easy to learn, easy to remember, and easy to use, which are key aspects for usable systems [1]–[3]. Brain activity as input modality also has potential in this area, as it can provide some insight in the intention of the user, without depending on external expression. Unfortunately, most of the current systems are still in the phase of proving that using brain activity for control is even possible, and are therefore not making full use of the intuitiveness this input could provide.

Manuscript received April 6, 2012; revised January 11, 2013 and May 10, 2013; accepted June 13, 2013. Date of publication October 1, 2013; date of current version November 18, 2013. This work was supported by the BrainGain Smart Mix Program of the Dutch Ministry of Economic Affairs and the Dutch Ministry of Education, Culture, and Science. This paper was recommended by Associate Editor I. Kotsia.

The authors are with the Human Media Interaction Group, Faculty of EEMCS, University of Twente, Enschede 7500AE, The Netherlands (e-mail: b.l.a.vandelaar@utwente.nl; d.plass-oudebos@utwente.nl; b.reuderink@utwente.nl; m.poel@utwente.nl; a.nijholt@ewi.utwente.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2013.2282279

A. Unreliable input

One thing physiology-based inputs have in common is that the interpretation of the input is often problematic. This is mainly because of the noisiness and ambiguity of the input, but also because of the problem of intentionality (see [4]–[6] for example). This noise and ambiguity will make an unreliable input channel. We try to clarify this through three examples.

1) *Keyboard:* In case of input through a keyboard, the keys that are pressed by a user are always recognized as what was typed in. There is (in case of a wired keyboard) no measurable noise between the keyboard and the computer and no problematic interpretation of the keystrokes. The only possibility for an error to occur is by uncontrolled motor activity of the user. A recognition or classification rate is not applicable in this case.

2) *Vision-Based:* In case of a gaze, gesture, or other vision-based interface: interfaces based on gaze, gestures, etc., depend on visually capturing (part of) the user. Either eyes or limbs point in a certain direction or make a certain movement which is classified by the computer and translated into an action within the system. Users can make errors by not accurately carrying out what they intended to do. As the system depends on the captured image, different kinds of artifacts and noise can hinder classification of the correct action. Pixel noise, different lighting, occlusion, movement of the person, movement of the camera, vibrations, etc., all may have a negative impact on the recognition rate. Input can change over time, for example as the user becomes more fatigued and is less expressive, or as the sun sets and the lighting changes. Besides, there can also be a large variability between users, such as between children and elderly, or men and women. As an example of ambiguity, when somebody waves their hand, it could mean good bye, hello, or even no. A final problem is intentionality (also referred to as the Midas

touch problem [7]). Not all actions will have a purposeful intention related to it. What if the user was not waving at the system, but waving to get rid of a mosquito passing by? Or in the case of an eye tracker, the user will already look at the system simply to take in information. In that case, not every eye gaze is meant as an input command. Especially when the system is always on, there will be times when the user is not purposefully interacting with the system.

3) *Brain:* Brain–computer interfaces (BCI) are dependent on the brain activity generated by the user performing a task. This can either be the user attending a visual, auditory or tactile stimulus, or by generating brain activity by actively performing a mental task (e.g., imaginary motor movement, performing mental calculus). Errors can occur when the user is not attending the stimulus or not (correctly) performing the mental task. But because of the highly varying nature of the resulting measured signal, the system will make errors in recognizing the users intention to attend/not attend or to perform the mental task or not. Recognition rates vary greatly between users and within users over time, largely independent of the type of BCI [8]. Related to the Midas touch problem, in BCIs it is difficult to detect whether the user is performing a task or not (e.g., idling) [9].

In the last two cases, there is an inherent noise (with different causes) on the input channel to the system. This noise makes recognizing the users intentions nontrivial. We therefore define unreliable input as the system not being able to reliably recognize the users intention, caused by inherent noise on the input channel.

B. BCI-Based input

In the case of brain–computer interfaces, each of these problems can be even more difficult to solve as there is no way to obtain a ground truth. Systems based on visual input can be evaluated using expert assessments on recordings of for example gaze and movements. For BCI based input this is very hard to do. First, BCIs require robust and noise-free brain activity recordings which is problematic at user's homes. For the general population, undergoing surgery to get electrodes implanted is also not a viable option, as the surgery, but also long-term implantation of these electrodes, are still too risky. Finally, there is the issue of response time. If a system is to be used for direct control, the response time should be minimal. This means that systems that depend on indirect measures such as increased blood flow in more active brain areas will be considered too slow for this purpose. What the general home-user then is left with is electroencephalography (EEG). Unfortunately, this measurement is highly sensitive to noise, both from the environment and from the user's body. It has a good temporal resolution, but because it uses electrodes on the outside of the head, it is difficult to only measure what we need to measure. Instead, a smeared out and attenuated signal from numerous interfering sources is measured. Even if all this would have been perfect, the brain is a very complex system, and specific areas may activate for different reasons. As an example, certain areas of the brain are involved in making gestures. These areas may also activate, however, when the user is imagining to make that gesture, or when the user is looking at somebody else making that gesture [10], [11]. The problem of intentionality also still remains, as the user's brain may be responding to something that is not at all related to this particular part of the interaction with the system. As a result, the interpretation of such physiological input modalities may never be perfect, and at the moment, brain input seems to be the least perfect of all.

In most studies concerning user control, the input itself is considered to be near perfect, although mistakes may still be made because the user is distracted, unskilled, or is unsure about what to do. The general solutions that are provided to solve problems caused by imperfect control are therefore generally to make sure that the system is responsive (short delays for feedback or updated system status), the feedback is easily understood and include an undo button [3]. There are very little guidelines for interfaces where the control input itself may be a critical issue. How many mistakes can be made before a system becomes unusable or unacceptable? One could argue that it is not the actual control that matters, only the user's perception, but how do these two relate? And do we even need to aim for perfect control? Especially in the case of entertainment applications, some imperfections may add to the challenge of the task up until the point the user gets frustrated. Keeping in mind an imperfect control when designing a game can keep the user in a state of flow while they deal with the problem and learn how to cope with the imperfect control [12]. In this paper, we let users utilize the keyboard for interaction and simulate unreliable input by manipulating the input from the keyboard and mapping it, with different levels of accuracy, to actions in the game.

II. Related Work

A. How to Measure Control?

There are many different measures of performance of the recognition by the system. Accuracy is the easiest one to understand as it is simply the percentage of correct interpretations. The opposite measure is the error rate. Precision (fraction of retrieved instances) and recall (fraction of relevant instances) are the de facto measures in the information retrieval and visual action recognition domain. However, when dealing with an interactive system the interval at which actions can be performed and speed of processing become important as well. Also, as accuracy and error are dependent on the number of classes and the ratio between samples from the different classes, more complex measures should be calculated.

For example, in the case of the interpretation of moving either left or right, a random choice should yield a 50% accuracy on average, but in the case of left, right, forward, and back, the random classifier would only achieve 25%. This means that performance of 75% accuracy in a two-class system is very different from the same accuracy in a four-class system.

For an example of acceptable accuracies within the BCI domain for a four-class problem: Ware *et al.* have evaluated the level of acceptable and desirable accuracy in a four-class steady-state visual evoked potential based brain-computer interface, with five participants, by incrementally decreasing the accuracy of the interface. They found that accuracy levels of at least 77% were accepted and desired [13]. These results are based on a very limited number of participants, but it may be an initial indication of the level of control toward which a BCI should aim.

However, in case of a two-class problem, if the user moves to the left 70% of the time, and the system would be a simple classifier that always selects the class with the highest prior probability (the class which has been used most in the past), it could already achieve an accuracy of 70%, where if the classes would be equal, the result would have been 50%. To address this problem, various performance measures have been designed. One of these is the receiver operating characteristic (ROC) which displays the ratio between the true positive rate (the fraction of correctly detected target movements) and the false positive rate (detecting a target movement). The related area under the curve (AUC) value of the ROC that gives the probability that target event (e.g., movement to the left) has a higher confidence than a nontarget event (all other events) [14]. While the ROC and AUC-ROC give reliable measure of performance when the prior probabilities of the events are unequal, they are inherently binary, i.e., they measure the performance for discerning only two classes. A performance measure for multiple classes that has been gaining popularity in the BCI field is the information transfer rate (ITR), which measures the amount of information expressed in bits that can be communicated through an unreliable channel per unit of time. In this case, the unreliable channel is the BCI, and the user is supposed to use an optimal encoding strategy for its message. Current, noninvasive BCIs have ITRs of up to 10–25 bits per minute [15]. The type of BCIs with the highest ITRs are so-called P300 speller systems. These systems make use of the event-related potential (ERP) associated to attending a stimulus in the EEG of the user. This ERP starts to show up 300 ms after the (visual) stimulus has been given to the user. By visually highlighting all letters of the alphabet the system can distill what letter the user was attending to [16].

The advantage of measuring performance with the ITR is that, when calculated based on mutual information (MUI) [17], it is insensitive to unequal prior event probabilities, and incorporates both the precision of and the time needed to detect an event. The more time you take for a selection, the more data the system can gather about the input you are trying to provide, the higher the resulting interpretation accuracy. But while a higher accuracy will increase the throughput, it will generally take more time which reduces the ITR in turn. Therefore this measure gives a good indication of the tradeoff between time and accuracy.

Because the ITR incorporates both the speed of communication, and the amount of information a single event contains, it might measure a quality that is as much related to the ability of the user to express its intent as possible with an objective measure. Similarly, the difficulty index ID in Fitts' law is also a measure of information, measured in bits [18]. The assumption of using an optimal encoding strategy made in the ITR might be difficult to achieve in practice however. Most BCI-based spelling applications do not use text prediction, and based on context the optimal predictions might change. But while an optimal encoding in the decoder might not always be feasible, some environments might be forgiving enough to let the user exploit alternative control strategies to optimize their information throughput. For example, an unreliable command to turn left might be replaced with turning right for a longer period when time and space permit. Finally, Furdea *et al.* proposed the written symbol rate for BCI based speller systems. This approach only accounts for correctly written symbols excluding corrections [19].

B. Dealing With Errors in BCIs

One example of how to deal with the Midas touch problem in BCI is the low-frequency asynchronous switch design, which is an added layer of control by which the BCI control can be turned on or off which improves overall classification rates [20].

In most cases, it is better for the system not to take action on an input than to take the wrong action. For example, in the case of a P300 speller, to correct an incorrect character, one selection is needed to delete it, and another selection to then select the correct one. Now, there are dynamic P300 spellers that do not make a selection until the confidence level for a specific option is above a certain threshold [21].

Such systems not only act when the certainty is high enough, but also use repeated inputs to increase the certainty for a specific selection. This last feature is part of many potential-based BCIs, as these features are very sensitive to noise and difficult to detect based on one repeat (single-trial). Systems based on other types of brain activity features could also make use of these same principles, by looking at the confidence levels of the classifier, or by combining successive classifications until a certain threshold is reached.

Quek *et al.* [22] developed a simulation tool for control in applications using a BCI but without the need of an actual participant and an EEG cap connected to the system. Various factors, such as error rate, nonstationarities due to a changing state of mind, noise, and delay, are incorporated in their models. However useful this paper is, for our study this approach was unusable for a couple of reasons. The study by Quek *et al.* [22] only simulates two classes (i.e., left and right), needs EEG data from users to generate a model, and would therefore be too time consuming and from a logistics point of view impossible for our intended experiment.

C. Perception of BCI Control

Various studies have been carried out not only focusing on accuracy, but also taking into account how users perceived exerting control through a BCI. In a comparison of user experience between actual and imaginary movement control of a BCI game, users had better control with actual movement, which also resulted in higher alertness. Imaginary movement was perceived as more challenging [23]. A more longitudinal study was done in which 14 participants played a BCI game repeatedly over a period of five weeks. They used three different mental task pairs during each session. The results indicate that the user preference for certain mental tasks is primarily based on the correct recognition of the tasks by the system, and second on the ease of task execution

[24]. To further improve possible recognition rates, Zander and Jatzev [33] developed and demonstrated an application detecting whether the errors occur on the user or system side. This could allow for new ways of dealing with errors, to improve the usability of any system [25]. Control and dealing with errors are important constructs in evaluating input methods that are unreliable. When evaluating games or virtual environments on user experience, which can be presence in virtual worlds [26] or user experience of BCIs [27] one has to not only look at the absolute performance, but also at the qualitative aspects of the input channel, such as the learning curve, obtrusiveness and intuitiveness. Combining improved recognition rates, user centered design, intelligently handling the inevitable errors, and proper user experience evaluation will all help in generating a better user experience for BCIs.

D. Perception of Control

From psychology, it is known that people tend to overestimate their ability to control events; this effect is called the illusion of control. In a laboratory study, participants had a varying control over a pair of lights. Even when their actual control was none at all, the participants indicated they had some level of control (the illusion of control) [28]. Langer demonstrated that this illusion of control is stronger when certain factors are present, such as competition, individual choice, familiarity with the action or elements part of the action, and level of involvement [29]. Thompson *et al.* [26] propose that this perception is already created simply from the intention to create a particular outcome (turn on the left light), and a possible connection between the action executed (press a button) and the outcome (left or right light on). If the expected outcome is positive, people tend to overestimate their level of control, whereas if the outcome is negative, people tend to downplay their amount of control [30]. Nijboer *et al.* [20] showed that mastery confidence was related to BCI performance. However, this was shown only in two out of four participants [31].

III. Methods

The goal of this paper is to investigate whether we can modulate control in a game and what relation exists between control and fun.

Hypothesis 1: If we can construct confusion matrices that resemble the various accuracy levels in the same way one experiences when controlling a BCI and this is validated by the perceived control of participants, then we have a proper way of modulating control.

Hypothesis 2: If the modulation of control is related to the reported values for fun, then we can further investigate what the relation is between these two and whether an optimum or point of no improvement exists.

Hence, we need to look at various conditions with different amounts of accuracy to evaluate the user experience. This requires many data points, not attainable by doing experiments in a lab. We therefore set up an experiment which can be done by using a web browser. The experiment consists of a

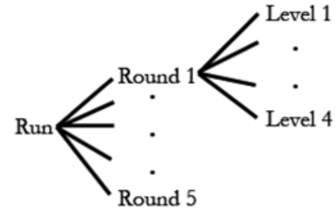

Fig. 1. Runs, rounds and levels.

Flash based game developed by a professional game studio. We varied the amount of control users had in the game and evaluated their user experience by administering a questionnaire after finishing the game. In the following sections we will elaborate on the design of the game, how we varied the amount of control and what data was used for analysis.

A. Experimental Design

The game consists of four levels, increasing in difficulty. Four levels in consecutive order make up one round, one or more rounds (usually five) make up one run. See Fig. 1 for an overview. Participants started every round in a randomly assigned condition. In total, there were 15 different conditions, each with a different amount of control (see Section III-C). Before starting each round users were given instructions on the control of the game, what they were supposed to do to finish the game as well as how to complete the questionnaire. After finishing a round, the in-game questionnaire was administered and a new condition randomly chosen for the next round. Participants in this experiment finished at least one round, and there was no maximum number of rounds. This means participants could opt to quit the experiment after every round. Participants were however encouraged to play for five runs. Five runs would amount to an average experiment duration of 25 min including breaks.

B. Game

In the game used for the experiment, participants controlled a hamster by pressing the four arrow keys on the keyboard. The game setting is an evil laboratory where experiments on hamsters are carried out on computer–brain interfacing with the goal to control hamsters. Users can take control over one hamster to lead it to freedom. A screenshot of the game can be seen in Fig. 2.

The amount of control of the user over the hamster varies randomly across rounds (Section III-C). One round in the game consists of four sequential levels, shaped in the context of the evil laboratory. A cage, a labyrinth, an office room, and the block, respectively, where the laboratory is situated have to be escaped from. Each level is a maze with dead ends and some occasional obstacles. Touching the obstacles causes the player to die and to be transported back to where they started in that level. When the user finishes a level, immediately the next level is presented. After the last level a questionnaire was presented (see Section III-E). After completing the questionnaire, the users could play another round, most probably with a different level of control. After five rounds the total amount of time they took to rescue five hamsters was recorded and

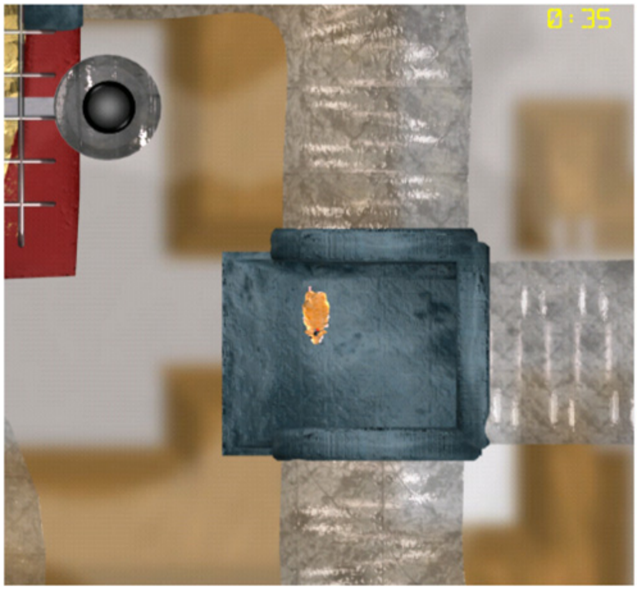

Fig. 2. Screenshot of the game that was used for the experiment.

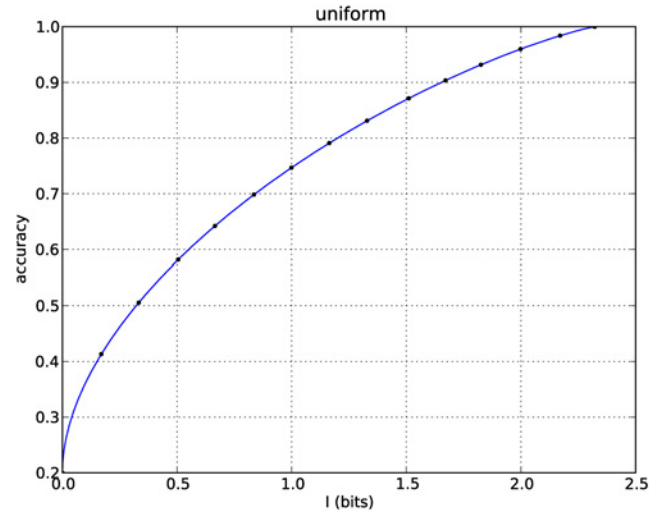

Fig. 3. Relation between the MUI and the accuracy for a five-class confusion matrix with the same probability for all correctly detected events, and the same probability of each pair of mistaking one event for another.

compared to the times that were already in the database to provide the user with a rank. This was supposed to motivate the participants to play again. If a certain round was too hard, the user had the opportunity to press a button skip round to skip the current round an go directly to the questionnaire. After the start of a level, it took 1 min for this button to become active.

C. Levels of Control

The level of control in the game was manipulated by using specific control schemes for each of the conditions consisting of different levels of control.

The events that are used to control the game consist of directives to move in one of four directions, or not to move at all, resulting in five possible events at each evaluation of the game loop. When the user has perfect control—which is usually the case with button-based input—each directive of the user is directly translated in the corresponding action. With unreliable controllers, such as a BCI, it is possible that a different, unintended action is performed. For each directive, action pair, the probability of making this mistake can be denoted, which results in a so-called confusion matrix. The behavior of an imperfect BCI with discrete output can be fully described by this confusion matrix.

The 15 conditions in the game are specified with a confusion matrix. We have chosen to start with the simple assumption that each class (four directives for movement, or no action) is detected correctly with the same probability (accuracy), and that all mistakes are equally likely

$$C_a = \begin{bmatrix} a & e & e & e & e \\ e & a & e & e & e \\ e & e & a & e & e \\ e & e & e & a & e \\ e & e & e & e & a \end{bmatrix} \tag{1}$$

where a is the percentage of correctly detected events, and $e = \frac{1-a}{5-1}$ is the rate for a specific confusion. Rows of C_a correspond to a specific directive (the ground truth) and sum to one, the columns correspond to a specific detection.

The specific accuracies are chosen such that the MUI of the confusion matrices C^a is distributed evenly with 15 points over the whole possible range. Given the rate of events, the ITR can be calculated from the MUI. The accuracies of the different throughputs with regular increases are displayed in Fig. 3. The logarithmic relation between the accuracy and MUI results in relatively few conditions with low accuracies. Given a information (I) of 0.0, the resulting accuracy will be 20% (Fig. 3) which equals a random distribution over five classes.

The MUI in bits is calculated as follows, with a discrete variable Y for the different directives, and discrete variable X for the different detections:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \frac{p(x,y)}{p_1(x)\, p_2(y)} \tag{2}$$

where $p(x, y) = C^a_{x,y}$ is the joint probability distribution function of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability distribution functions of X and Y, respectively. Note that the marginal probability distributions are assumed to be uniform. This is the same assumption used in [15] for ITR calculation.

In summary, we constructed 15 levels of control based on the equally spaced MUI of control schemes with equal probabilities of correct detection, and equal probabilities for each confusion.

D. Participants

Participants to the experiments were invited to play the game through various means (e-mails, social media, and mouth-to-mouth). Based on IP addresses, 200 unique participants started a round. In total, 351 rounds were started. Two hundred and twelve (60.4%) of these runs were continued through the four levels and properly filled in the questionnaire. The average age of all participants was 24.99 years (SD = 7.76). Eight (3.8%) participants refused or gave a bogus

TABLE I
CONSTRUCTS AND ITEMS

Construct	Items
Control	1. I had the feeling the hamster did what I wanted it to do 2. I had the feeling the computer was following my commands
Frustration	3. I was frustrated while playing the game
Fun	4. Playing the game was fun
Empowerment	5. During the game: I felt empowered
Satisfaction	6. During the game: I felt satisfied

answer on the age question. 52.5% of the participants indicated male and 47.5% of the participants indicated female on the gender question. Only 45.8% of all participants replied to this question. Twelve participants played the game for the desired total of five rounds and got a high score ranking in the game. The fastest to finish did this in a time of 21 min and 42 s, the slowest in 35 min and 3 s.

E. Questionnaire

The questionnaire was presented to the user within the Flash game, before they could continue with another run in the game. The questionnaire included three open questions and six visual analog scale (VAS) items. Two open questions were included to gather basic demographics, namely age and gender. The third open answer box was for general remarks and additions.

The six VAS items measured the amount of fun, engagement, and control the users experienced in the game. Table I shows the items for each construct. The way users answered the VAS was through a graphical slider. Of course, the experiment being carried out by means of a computer this VAS was digital, but approaching the analogue domain with values ranging from 0 to 100. One click on the scale put the slider on the designated point, also indicated by the changing number right below the scale. Users could correct their answers until they clicked on the next button, to go to the next page with three items.

IV. RESULTS

In this section, we will analyze the results from the questionnaire data. First, we will report the data on the perceived control to assert that our method of varying the amount of control is also experienced by the user in the correct way. In the section thereafter, we will analyze the relationship between the amount of control induced by the confusion matrices on the amount of fun the user experiences.

A. Perceived Control

First, we cleaned up the data from obviously erroneous responses. If one of the following criteria were met the response would removed from our data set.

1) Partly filled-in questionnaires: if only the first or the first and second page with questions was filled in and not the subsequent page(s), the complete response was removed from the data set.
2) Meaningless questionnaires: if all questions were answered with the initialization value (the VAS scale was initialized in the middle with a pointer), the complete response was removed from the data set.

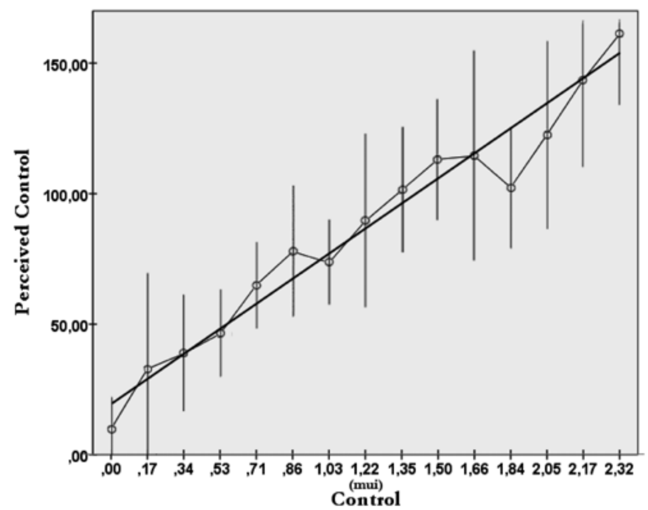

Fig. 4. Median of perceived control including error bars and linear regression of control (x-axis) versus perceived control (y-axis).

In Section III-B, we mentioned that participants had the ability to skip a round after 1 min. We included these records as it still provides us with useful data. After the cleanup process, we got 158 data points to base our analysis on. Looking at the cleaned questionnaire data we first describe the distribution of data points over the conditions. We expected an even distribution of data points over the conditions as these were randomly initialized. The minimum amount of data points per condition was seven for the lower half of the accuracy scale and ten for the higher part of the scale.

Then, we constructed the scale perceived control. This scale was made up out of two items, "I had the feeling the hamster did what I wanted it to do" and "I had the feeling the computer was following my commands." A reliability analysis of the proposed scale resulted in a Cronbach's Alpha of 0.885 which made this scale a proper measurement of how much control the participants experienced in the game. To validate if we indeed varied the amount of control as we intended a linear regression was carried out. This revealed a significant linear trend, explaining 50% of the variance in the data ($R^2 = 0.499$, $p < 0.001$) as can be seen in Fig. 4.

Another item, frustration, further validates our method of influencing control. Linear regression showed a significant negative trend ($R^2 = 0.133$, $p < 0.001$), as shown in Fig. 5.

B. Fun-Control

To validate our most important hypothesis, control positively influences fun, up to a certain point where fun is decreasing because the challenge is becoming less, we analyze the questionnaire data for fun and compare it to the amount of control. To assess whether a linear model or another model would explain our data better, we first performed a linear regression. This showed a significant trend ($R^2 = 0.291$, $p < 0.001$) through the data. Second, third and higher order polynomials were also tested. The third-order polynomial, which can be seen alongside the first-order (linear) polynomial in Fig. 6,

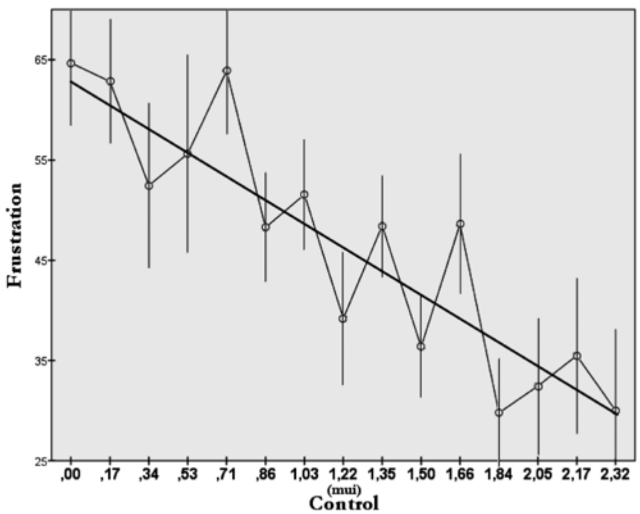

Fig. 5. Median of frustration including error bars and linear regression of control (x-axis) versus frustration (y-axis).

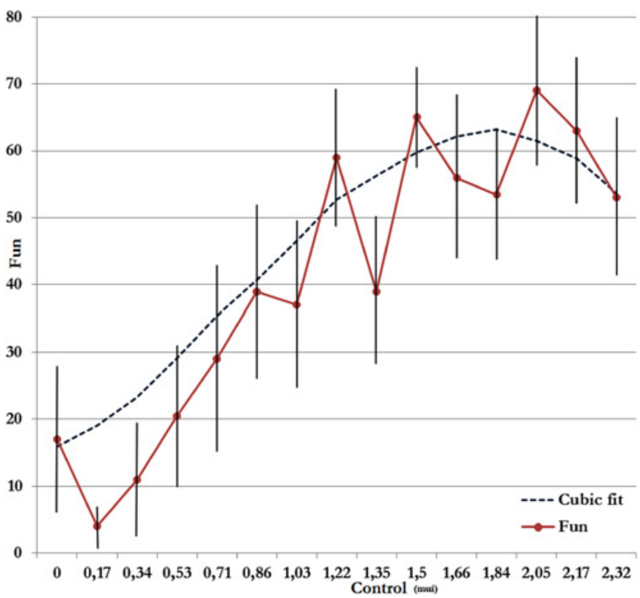

Fig. 6. Median of fun (y-axis) including error bars versus control (x-axis) and fourth-order curve fit.

proved to provide the best fit with an explained variance in the data of 34.9% ($R^2 = 0.349$, $p < 0.001$), taking into account that higher order polynomials yield a slightly higher R^2 but also require sacrificing another degree of freedom for every incrementation of the order. Another measure for this that takes the added complexity of higher order polynomials into account is the Akaikes information criterion (AIC) [32]. The gain of less information loss is resembled in a lower AIC. While the difference in the AIC for the second- and third-order polynomial is substantial, the difference between the third and fourth orders is only marginal compared to the former.

The third-order polynomial and the medians of the data points show a clear downward tendency from the linear trend after the mui=2.05/acc=96.73% point. This supports our hypothesis that control is needed for fun, up to a certain point, after which fun decreases. Another interesting fact is that the condition in which movement is completely random

TABLE II

FIRST THROUGH FOURTH-ORDER MODELS WITH RESPECTIVE R^2, AIC, AND THE CHANGE IN AIC FROM THE LOWER ORDER POLYNOMIAL

Models for curve fitting			
Model	R^2	AIC	delta(AIC)
1th order (linear)	0.291	977.683	n/a
2nd order (quadratic)	0.336	969.434	-8.249
3rd order (cubic)	0.349	968.305	-1.129
4th order	0.356	968.170	-0.135

(20% accuracy, mui=0) participants apparently found it relatively fun to play the game. When participants do have some amount of control in the condition following it (mui=0.17/41,22% accuracy) this effect is gone.

V. DISCUSSION AND CONCLUSION

This paper showed that it is possible to influence the amount of perceived control the user has in a game that has a four-class, 2-D control of navigation. This is supported by the fact that users reporting the amount of control they experienced showed a strong relation to the amount of control they were given. We hypothesized an illusion of control; an overestimation of control on the near perfect side of control. Our data did however not support this hypothesis. We also did not find support for an underestimation of control at the low side. A possible explanation for this might be that in our study, users are given a certain amount of control defined by the confusion matrix in a session. This given amount of control is something they cannot alter by increasing their effort. Whereas with a BCI user may alter their amount of control, to a certain extent by the effort they put in. In this case, the effort put in by the user is highly dependent on the motivation they have, which may be related to the performance of a P300-based BCI [33].

The second part of our analysis showed that the amount of control largely explains the amount of fun one experiences. Although fun is dependent on the amount of control, at a certain point an optimum is reached. Our analysis showed that in our experiment after 96.73% of accuracy the fun decreases. This could be explained by the concept of flow [34], where challenge is related to skills in which an optimum exists where a state of flow is achieved. In this state of flow the skills of the user and the challenge asked from the user are both high and in balance. If the skill increases, the user would shift into boredom, if on the other hand the challenge increases, the user shifts into anxiety or frustration. This is related to the results we see in our data. Users have a certain skill of controlling the hamster and have a varying amount of control (the challenge), up to a certain point this challenge is more suited to their skill, until the optimum is reached. After this the fun decreases and general user experience shifts into boredom. This explanation is also supported by some user comments, for example, “This is the third time I played this game, and the hamster listened quite well. Especially if your first hamster were never obeyed anything [sic], a well-listening hamster is almost boring.” This mechanism of shifting toward boredom could therefore be the reason the amount of fun participants experienced is decreasing after the optimum. At the other end of the scale, the concept still holds: users reported being frustrated through the respective item in the questionnaire, as well as through the open question while playing the game with a low amount of control, for example, “I haven’t quite forgiven hamsters after the last game even though this one was better” and “Frustration thy name is hamster. You know he was perfectly happy in his cage...” This is also what is often seen in BCIs, if the recognition accuracy is too low, the BCI is just a frustration to the user. It might give the user false hope that leads to frustration.

As in almost all of HCI research, the results in this paper, especially the point for the highest amount of fun, are based on just this game with which we used to experiment. Using another game would probably yield another optimum at a different amount of control. Probably even using other level layouts would alter this result. However, we show that this effect is apparent in this particular setting. The game mechanics also included obstacle-like enemies in the harder levels, which upon collision would send you back to the beginning of the level (resetting one’s progress). This means our simulation also includes the effects of making an error and users experiencing the cost of an error when using a real BCI. Without the possibility of making errors that have such a high cost, the challenge would probably be smaller as one will eventually get to the end of the level. Trying to anticipate on erroneous movements (and steering clear of the obstacles) is part of the challenge in the last two levels. A smaller challenge with the same amount of control could decrease frustration for the low accuracy conditions and increase boredom (and decrease fun) for the high accuracy conditions. We think that therefore the optimum could shift to the left in such a game.

Still, the game can be quite challenging, even with perfect control, when there is a motivation to finish as quickly as possible. In games that are simpler, for example, that only use two classes for navigation, or are less challenging in game play, the optimum probably shifts to the lower end of the control scale. Games with some kind of BCI control often only have one dimension of control, so the previous is especially applicable to these kinds of games.

Our results show that in the lower levels of control, the accuracy has an important influence on the user experience. While a higher accuracy leads to a better user experience in the lower parts, the better the accuracy gets, the less influence it is going to have on the user experience. In this particular game, we estimated this critical point to be around 96.73% (although around 86% one can already see the effect in the data). In comparison to state-of-the-art BCIs this is a very high accuracy. There are, however, some tradeoffs that have to be made in designing and implementing a BCI. The tradeoff between speed and accuracy is an important issue for this discussion. The longer the BCI gets to measure the EEG of the respective mental task, the higher the chance it will make the right decision. However, this sacrifices speed. Another common tradeoff is the number of classes used in a BCI. While we tried to simulate a five-class (four directions + no operation) often BCIs are still implemented with a two- or three-class operation that results in a less complex game with either a 1-D control or a timed sequential control in which a direction to walk toward has to be chosen based on a timer. In the first case, the accuracy of classified results will be higher than in a four/five-class interaction, but the richness of the interaction will be lower. This could mean that the point where either an optimum is reached or plateauing sets in will be lower than the point we found in our study. Therefore, it might not be worthwhile striving for the highest possible accuracy at the cost of other factors that might have a bigger influence on the user experience. In the second case, the richness of interaction might be higher because of more possible actions, but the sequential selection of commands could increase the latency of the interaction to a point where it might be just a frustration for the user.

Overall, in our results, the exact point at which added accuracy does not improve the user experience anymore is not that important; however, the finding that there is such a point at all, is. We think these findings could be an advocacy for not only optimizing for accuracy but taking other factors into account as well. Even outside the field of BCI this finding could be useful. Essentially in all modes of interaction where uncertainty is a factor, like physiological computing or gesture-based interfaces a tradeoff exists between system response time and accuracy. Although recognition accuracies might be higher or lower and interaction options might be more diverse than in a typical BCI, the effect we found might be present in these interaction methods as well.

Until now, no perfect BCI has existed. However, according to our findings a perfect BCI controlled game might not be a system that is only optimized for accuracy. At some point accuracy does not influence the fun anymore and other factors start to become more important. Therefore, turning the shortcomings of a BCI into a challenge for the user, a challenge outside of the game itself, might be a possible way to create a fun game.

REFERENCES

[1] C. Mohs, J. Hurtienne, J. Israel, A. Naumann, M. Kindsmüller, H. Meyer, and A. Pohlmeyer, “IUUI: Intuitive use of user interfaces,” *Usability Professionals*, vol. 6, pp. 130–133, 2006.

[2] A. L. Blackler, V. Popovic, and D. P. Mahar, “Toward a design methodology for applying intuitive interaction,” in *Proc. WonderGround 2006 Design Research Soc. Int. Conf.*

[3] J. Nielsen, *Usability Engineering*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[4] L. Deng and X. Huang, “Challenges in adopting speech recognition,” *Commun. ACM*, vol. 47, no. 1, pp. 69–75, 2004.

[5] R. Poppe, “A survey on vision-based human action recognition,” *Image Vision Comput.*, vol. 28, no. 6, pp. 976–990, 2010.

[6] R. Jacob and K. Karn, “Eye tracking in human-computer interaction and usability research: Ready to deliver the promises,” *Mind*, vol. 2, no. 3, pp. 573–605, 2003.

[7] R. J. K. Jacob, “Eye movement-based human-computer interaction techniques: Toward noncommand interfaces,” *Advances Human-Comput. Interaction*, no. 4, pp. 151–190, 1993.

[8] C. Vidaurre and B. Blankertz, “Toward a cure for BCI illiteracy,” *Brain Topography*, vol. 23, no. 2, pp. 194–198, 2010.

[9] D. Zhang, Y. Wang, X. Gao, B. Hong, and S. Gao, “An algorithm for idle-state detection in motor-imagery-based brain-computer interface,” *Intell. Neurosci.*, vol. 2007, p. 5, Jan. 2007.

[10] L. Fadiga, L. Craighero, and E. Olivier, "Human motor cortex excitability during the perception of others action," *Current Opin. Neurobiol.*, vol. 15, no. 2, pp. 213–218, 2005.

[11] D. McFarland, L. Miner, T. Vaughan, and J. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography*, vol. 12, no. 3, pp. 177–186, 2000.

[12] A. Nijholt, D. Plass-Oude Bos, and B. Reuderink, "Turning shortcomings into challenges: Brain-computer interfaces for games," *Entertainment Comput.*, vol. 1, no. 2, pp. 85–94, Oct. 2009.

[13] M. Ware, P. McCullagh, A. McRoberts, G. Lightbody, C. Nugent, G. McAllister, M. Mulvenna, E. Thomson, and S. Martin, "Contrasting levels of accuracy in command interaction sequences for a domestic brain-computer interface using SSVEP," in *Proc. 5th CIBEC*, 2010, pp. 150–153.

[14] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2005.

[15] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[16] L. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, Jul. 2006.

[18] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.

[19] A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kbler, "An auditory oddball (p300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.

[20] G. Birch, Z. Bozorgzadeh, and S. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials," *IEEE Trans. Neural Syst. Rehabilitat. Eng.*, vol. 10, no. 4, pp. 219–224, Dec. 2002.

[21] H. Serby, E. Yom-Tov, and G. Inbar, "An improved p300-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabilitat. Eng.*, vol. 13, no. 1, pp. 89–98, Mar. 2005.

[22] M. Quek, D. Boland, J. Williamson, R. Murray-Smith, M. Tavella, S. Perdikis, M. Schreuder, and M. Tangermann, "Simulating the feel of brain-computer interfaces for design, development and social interaction," in *Proc. CHI*, May 2011, pp. 25–28.

[23] B. van de Laar, B. Reuderink, D. Plass-Oude Bos, and D. Heylen, "Evaluating user experience of actual and imagined movement in BCI gaming," *Int. J. Gaming Comput. Mediated Simulations*, vol. 2, no. 4, pp. 33–47, 2010.

[24] D. Plass-Oude Bos, M. Poel, and A. Nijholt, "A study in user-centered design and evaluation of mental tasks for BCI," in *Proc. 17th Int. Multimedia Modeling Conf.*, vol. 6524. Taipei, Taiwan, Jan. 2011, pp. 122–134.

[25] T. Zander and S. Jatzev, "Detecting affective covert user states with passive brain-computer interfaces," in *Proc. 3rd ACII*, 2009, pp. 1–9.

[26] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence Teleoperators Virtual Environments*, vol. 7, no. 3, pp. 225–240, 1998.

[27] B. L. A. van de Laar, H. Gürkök, D. Plass-Oude Bos, F. Nijboer, and A. Nijholt, "Perspectives on user experience evaluation of brain-computer interfaces," in *Proc. 6th UAHCI*, LNCS 6766. Jun. 2011, pp. 600–609.

[28] L. Allan and H. Jenkins, "The judgment of contingency and the nature of the response alternatives," *Canad. J. Psychol./Revue Canad. Psychol.*, vol. 34, no. 1, pp. 1–11, 1980.

[29] E. Langer, "The illusion of control," *J. Personality Social Psychol.*, vol. 32, no. 2, pp. 311–328, 1975.

[30] S. Thompson, W. Armstrong, and C. Thomas, "Illusions of control, underestimations, and accuracy: A control heuristic explanation," *Psychol. Bull.*, vol. 123, no. 2, p. 143, 1998.

[31] F. Nijboer, N. Birbaumer, and A. Kübler, "The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis: A longitudinal study," *Frontiers Neuroprosthetics*, vol. 4, pp. 1–13, Jul. 2010.

[32] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inform. Theory*, 1973, pp. 267–281.

[33] S. Kleih, F. Nijboer, S. Halder, and A. Kübler, "Motivation modulates the p300 amplitude during brain computer interface use," *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1023–1031, 2010.

[34] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, NY, USA: Harper and Row, 1990.

Authors' photographs and biographies not available at the time of publication.