Design and Study of Two Applications Controlled by a Brain-Computer Interface Exploiting Steady-State Somatosensory-Evoked Potentials

Jimmy Petit, Jose Rouillard, Francois Cabestaing

Design and Study of Two Applications Controlled by a Brain-Computer Interface Exploiting Steady-State Somatosensory-Evoked Potentials

Jimmy Petit [1], José Rouillard [1] and François Cabestaing [1]

[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL,

F-59000 Lille, France

ABSTRACT

Brain-Computer Interfaces (BCI) allow users to interact with machines without requiring muscular activity. Thus, patients with heavy motor impairment can benefit from these systems. We have implemented an electroencephalography-based BCI which provides four distinct commands. Our system exploits the Motor Imagery of the subject and four different states of mind: imagination of a movement with the left or right arms, both arms simultaneously and no imagination at all. In addition, the BCI exploits specific neurological markers called Steady-State Somatosensory-Evoked Potentials. They are vibrating devices taped on the user's wrists. These markers are measurable on the cortex using electroencephalography. This paper focuses on the Computer Human Interaction aspects. We describe the design and study of two applications controlled by this BCI. The applications differ in two characteristics: their inertia, or rhythm of information flow perceived by the user, and the "punitiveness" of the application in case of mistakes.

To study the user experience in perfectly controlled conditions, we used a so-called "sham" feedback in the BCI loop rather than real feedback computed by analysing the user's brain waves. With sham feedback, the BCI provides commands with an *a priori* defined accuracy. We performed a user experiment of the two applications over a group of ten healthy participants. They tested both applications for different sham accuracies, varying from 45% to 90%. This permits the study and modelling of the relationship between the perceived usability of the system and the performance of the BCI.

INTRODUCTION

We implemented a Motor Imagery based BCI which provides four commands. We conducted a user experiment to evaluate our applications over a group of ten healthy subjects. This user experience is the last session of a four-session-long experiment (ethical comity of Lille University, reference: 2020-417-S81). The other three sessions are more focused on studying the neurophysiological aspects of the BCI. However, describing these sessions and their results is out of the scope of this paper dedicated to the Computer-Human Interaction aspect.

For the user experiment, we have developed two applications. However, BCI using Motor Imagery and four different commands tends to have a low accuracy, which greatly improves with a lot of training. Since the users had no time to train themselves, we used of sham feedback to simulate various levels of performance (Park, et al., 2021). The sham feedback has various advantages compared to testing the real performances of the system:

1. The classifier used by the BCI has already been trained before the user experiment session. However, within one or two weeks between the two sessions, the mental activity would have drastically changed, and it would be likely that the classifier would produce the same output to any given input, resulting in a succession of commands like Turn Left, Turn Left, Turn Left, … Sham feedbacks avoid locking the user in that situation.
2. A consequence of implementing sham feedback is that we must set the desired command (a "Good" action) for any given state of the application using a combination of level design and instructions. This allows us to then perform an offline relevant test of different classification algorithms and build on a dataset where the number of commands is reasonably balanced. For example, without sham feedback, the subject could be stuck in a state where it must use the same command repeatedly, which seriously unbalances data set.
3. For the user experiment, we use various levels of sham, ranging from 45% to 90%, which allows us to study the relationship between the system accuracy and various measured aspects in the questionnaires, which is not possible otherwise.

The objective of this article is to provide the details of the applications and how we encourage the user to use a balanced number of commands. In addition, we study the effect of the applications and the user experiment on the users themselves using questionnaires investigating fatigue, mood orientation, or mental workload, for example. Finally, we study the correlation between the perceived usability of the system, assessed with a System Usability Scale, and the performance of the system. Therefore, for a given BCI performance, we can predict what mean degree of perceived usability will be achieved, and *vice versa*.

APPLICATIONS DESIGN

In this section, we will endeavour to present in detail the applications, to allow their reproducibility. In addition, as our user experiment uses sham feedback, we are going to introduce our definition of "Good" and "Bad" commands in the applications.

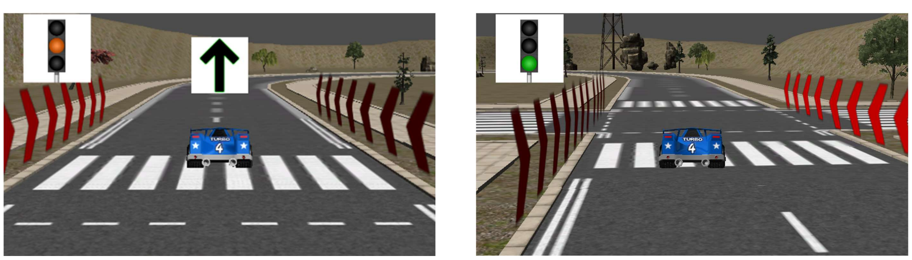
Description

The first application is a kart-driving application. The kart is controlled with the four available commands provided by the BCI: Move Forward (MF), Turn Left (TL), Turn Right (TR), and Do

Nothing (DN). To balance the number of "active" commands, *i.e.*, MF, TL and TR, the kart moves on an eight-shaped road. During a left turn of the road the user is instructed to perform a TL command, and *vice-versa*. In a straight part, the user is instructed to perform the MF command. Instructions are given using a sheet at the beginning of the experiment. A Turning command performs two actions on the kart: the vehicle turns and receives an additional pushing force. Therefore, one TL or TR command will make the kart follow a curved trajectory. To help stay on tracks, semi-transparent walls, with red arrows on them, follow the kart and apply a repulsive force to it as soon as it gets too close to them. Figures 1 shows two screenshots of the application. Figure 1 (a) shows the feedback given to the user: the kart moves or gains speed and a image is displayed on top of the kart. The kart was on a straight line, therefore, it has a fixed percentage, like 60%, of doing MF according to the sham feedback. If a "Bad" action must be performed, then the action is randomly and uniformly selected between TL, TR, and DN.

The second application is a puzzle-solving type of game called SokoBCI, inspired by the game Sokoban (Hiroyuki Imabayashi, 1982). The user controls the movement of a 3D avatar who must plant trees. The user must solve the levels using the smallest number of commands. To balance the number of active commands, one run of the application is composed of 2 levels. The first level is a simple symmetric level that forces the user to use an uneven number of TL and TR commands. During the solving of the first level, the system counts the number of TL and TR commands. Level 2 is chosen to encourage the user to use the command least used during level 1. The actual vision from the user's point of view during level 2 is displayed in figure 2. The level designs and the aforementioned process are displayed in figure 3, which is a top view of the levels. When the user plants a tree, the tree appears on the red square and the red square turns green. During a run of the application, as soon as the user finishes level 1, the appropriate level 2 appears on the screen and the loop of commands starts again. To give time to the user to think about the best solution, the first command of level 2 is completely ignored by the computer, during a dozen seconds of break.

For both applications, a three-colour light is displayed at the top-left corner of the screen. The green light gives the cue to the user to start the Motor Imagery task to send a command to the BCI. After six seconds, the light turns to orange, indicating the feedback period, during which the kart or the avatar will react accordingly to the given command. After three seconds of feedback, a short break with random duration, between three to four seconds, is given to the subject with the red light.



(a) *(b)*

Figure 1: Screenshots of the kart (a) on a straight line (b) at the cross section of the eight shaped road.

Definition of "Good" or "Bad" Command: Specific Case of SokoBCI

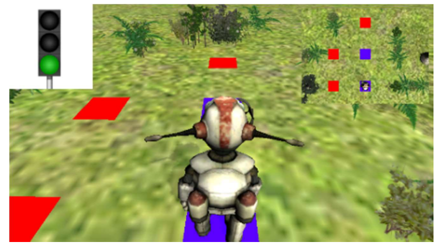


Figure 2: User view at the beginning of a level 2 (Turn Left induced-design).

The description of the Kart driving task is quite clear about the instructions given to the subject. We believe that the definition of "Good" or "Bad" command is straightforward in this case. However, for the SokoBCI it is less clear. In our user experiment, the user is instructed to "use the least number of commands to plant all the trees. If the system does an unwanted action, [example given], you might need to change your initial plan.". When observing the level design, one can observe that the solution for level one is to rotate once; left or right, and plant a tree, then rotate twice, in the same direction, to plant the second one. In this situation, the commands TL and TR are considered equally "Good" commands, the commands MF and DN, will have the same effect which is leaving the avatar motionless and are "Bad". However, during the U-turn, the initial rotation will dictate the next choice of the user, the first rotation will be either an error or not, depending on the user's mental decision. Given the instruction, the user will adjust their mental plan to pursue the rotation, end therefore the "Good" action becomes predictable.

In the first level, the avatar can face four different directions and the tree planting can have four different states (no tree planted, the left tree only, the right tree only, and both trees planted): there are thus twelve (4 * (4 - 1)) different possible game states, the game state where all trees being planted being naturally ignored. Four of the game states have two "Good" actions. In the second level, left or right, the avatar can face four different directions, through two different locations, and the tree plantation can have eight different states (2^3 possible situations). Therefore, we have fifty-six (4 * 2 * (2^3 - 1)) possible states of the game, of which thirteen states have two simultaneously "Good" actions.

Inertia and "Punitiveness"

The first difference is in inertia. The kart continues to move forward during the green light period. It is a vehicle that gains momentum by receiving the TL, TR, and MF commands. The user must analyse the road and speed to decide when to send the next command, even during the red-light periods which represent the breaks. The second difference is the punishment for mistakes, which we call "punitiveness". In the kart application, the semi-transparent walls follow the kart as it moves along the track, pushing the kart in the opposite direction as it gets closer. In this case, even mistakes tend to move the kart forward, however slowly. In the SokoBCI, on the other hand, a mistake can be much more frustrating as the user will have to make a correction move in most cases.

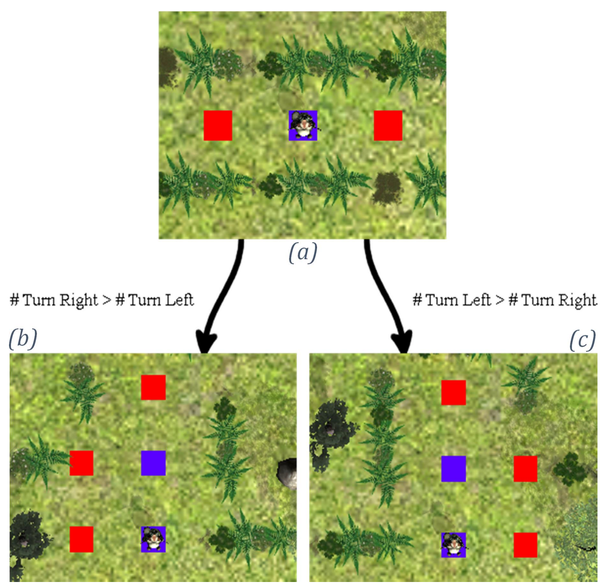


Figure 3: Top view of the SokoBCI levels. (a) level 1. Depending on the number of TL or TR used to solve the level 1, one level 2 is chosen (b) or (c). The blue squares show the possible location of the avatar, accessible with the Forward command while facing the proper direction. The red squares show the targets. The user has to orient the avatar towards a target and use the command MF to plant a tree.

TEST OF THE APPLICATIONS

We conducted a user experiment with a group of ten healthy participants, seven males and three females. The average age of the participants is 23.8 years (std: 3,2 years, min: 19, max: 28). In this section, we to present the experimental protocol.

Protocol

The subject sits in front of the computer and fills out a pre-session questionnaire. Then, the experimenter gives an instruction sheet to the subject, who reads it and can ask any questions needed for good comprehension. Afterwards, the experimenter installs the brain computer interface device on the user, as well as the devices that delivers the mechanical stimulation to the wrists. After this installation step, the proper tests can begin. We start by demonstrating one of the applications, chosen with a pseudo-random order across subjects, we remind the user the objective of the application and the instructions. After that, the user performs four blocks of recording with the application. Each block has a specific accuracy for the sham feedback, the four chosen accuracies are 45%, 60%, 75%, and 90%. The order is also pseudo-random. After four blocks of recordings, the demonstration and test steps are repeated for the other application. The subject has a mandatory break of 3 minutes (minimum) between each recording block.

The subject fills out questionnaires after each block and the experimenter conducts a debriefing interview with the subject at the end of the experiment. The experimenter asks the subject to fill out the questionnaire independently from one block to another. We have three different questionnaires. The first one is about behaviour measurement, the second one is a NASA Task Load IndeX without the pairwise comparison of the questionnaire's dimensions, called Raw TLX (Hart, 2006), and the last one is the System Usability Scale (Brooke, 1996) (Brooke, 2013), or SUS. The last two are given to be completed at the end of each block. The first questionnaire is given as a pre-session questionnaire and after the first block, the fourth block (*i.e.*, four blocks of the first application), the fifth block and the eighth block (*i.e.*, four blocks of the second application), to measure changes in behavioural data during each application.

Table 1: Interview guide for the experimenter

1)	What did you think of the experiment?
2)	What did you think about the difficulty to adjust your mental plan to the system's error during the SokoBCI?
3)	Inform the subject about the sham feedback: what is it about, how is it done, and why we do it.
4)	Did you ever suspect the sham feedback? [If yes, when?]
5)	Did you suspect different levels of performance/accuracy?
6)	Did you discuss about performances during the previous session [which also had sham feedback] with your friend/colleague? [Question optional, question asked only if the friend/colleague also participates in the experiment.]

The Behavioural questionnaire contains four items assessing "Awakeness", Tiredness, Mood Orientation, and Emotion Intensity with a five-point Likert scale. Additionally, a space for comments is left for subjects to express themselves on the application they have just tried. The Raw TLX aims at measuring the mental workload of the application with six items: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration measured on a twenty-point Likert scale. The SUS aims at measuring the usability of the system with a ten-item long questionnaire, also using a five-point Likert scale.

In the debriefing interview at the end of the session, we discuss with the subject different aspects of the whole experiment. Table 1 presents the different questions asked, and their order. At the end of the session, some subjects had strong time constraints, which led us to shorten some interviews.

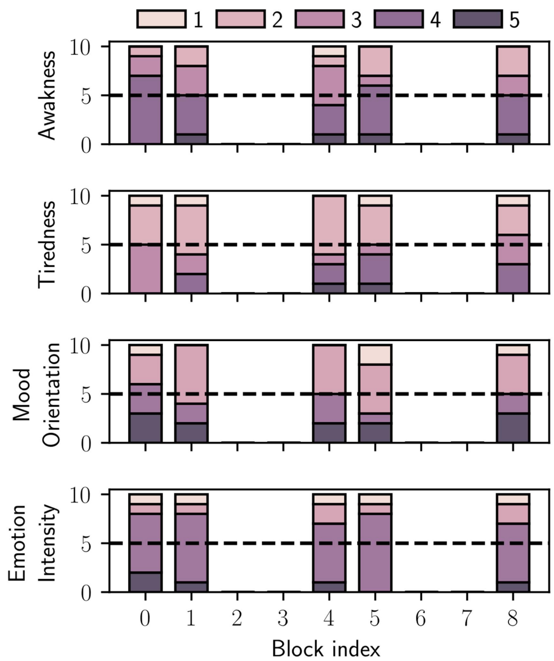


Figure 4: Stacked bar plots showing the distribution and evolution of Behavioural data over the subjects.

Behavioral Data - Results

Results of the Behavioural data analysis are displayed in figure 4. We can observe that, "Awakeness", Mood orientation (positive to negative), or Emotion Intensity do not seem to evolve during the session. However, Tiredness tends to increase, indeed the majority mention (the mention that crosses the 50% threshold) became three (out of five) by the end of the session. A few subjects filled out the mention number four during the session too, which they did not at the beginning of the session. To conclude in a few words, the eight blocks of recording seem to cause a little fatigue to our subjects and additionally, the three other dimensions do not evolve much.

The Raw TLX results are presented in figure 5. When observing the distribution of the Raw TLX score against the sham accuracy, no trends seem to appear. No difference can be observed between applications either. However, the individual dimensions of Performance and Frustration are highly inversely proportional to the sham accuracy. It seems to hold for the Effort dimension of the questionnaire too, however to a lesser extent than the previous two dimensions. The subjects seem to have well experienced the difference in the BCI performances and the frustration level diminished as the performance increased, likely caused by the decreasing number of mistakes performed by the system. Interestingly, as the accuracy increases, we measure a difference in Mental Demand between the applications. With sham accuracy at and above 75%, the subjects seem to experience a Kart application more mentally demanding than the SokoBCI application. This is confirmed using a dependent samples *t*-tests[1]: at 75% a p-value of 0.0012 (< 0.05) and at 90% a p-value of 0.019 (< 0.05) are computed. At 75%, the mean difference in Mental Demand between the Kart and the SokoBCI is at 2.2 points (std: 1.6) while being at 2.5 points (std: 3.1) at 90% of sham accuracy. This difference can be explained by the difference in inertia between the

[1] We failed to reject the hypothesis of normality of the four samples using Shapiro-Wilk tests and a rejection threshold of 5%, therefore we assumed the normality of the samples.

two applications. In the kart, when the three-colour light is red, the kart is still moving, and the user must think to anticipate the movement of the kart. Additionally, the effect does not occur to low accuracy as multiple errors tend to slow down, or even stop, the kart. Finally, a weak difference between the application in the Frustration level is also observable for the lowest accuracy as the mean and median are systematically higher in the SokoBCI than in the Kart. This could be explained by the "punitiveness" of the SokoBCI, an error is much more costly in this application than in the Kart-driving one.

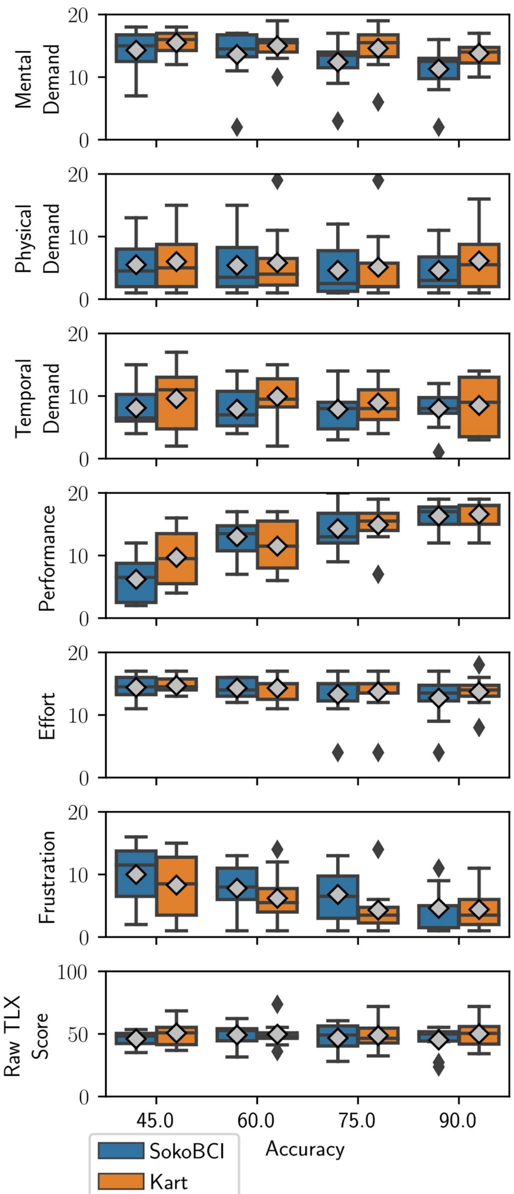


Figure 5: Box plots of the Raw TLX results, grouped by the application type. The six first lines show the individual dimension of the questionnaire, while the seventh line show the score of it.

Relationship Between Performance and SUS

Figure 6 shows the results of the SUS questionnaires. Firstly, the SUS scores are positively correlated to the sham accuracy.

According to the slope coefficient of linear regression, the effect is stronger in the SokoBCI application than in the Kart. It can be also explained by the higher "punitiveness" of an error in SokoBCI than in the Kart.

Bangor *et al.* added an adjective rating scale to mean SUS scores, see figure 7 (Bangor, et al., 2009). In doing so they also provide an interpretation of SUS scores using an adjective, school grading system, or acceptability range from previous works. According to their results, a system would start to be acceptable with a SUS of 70 and everything below had usability issues and is cause for concern. Therefore, using these models, in a BCI with similar features to the SokoBCI, the system would become acceptable with a minimum performance level of 85.4%, whereas a system like the Kart application would need a minimum performance level of 74.3%.

Questionnaire Comments and Debriefing Interview Results

Reading the comments and notes during the experimenter's debriefing allowed us to identify four categories of feedback: **Strategy**, the strategy found by the subject to perform the task; **Application Comparison**, the subject expresses some preference toward one application over the other; **Positive Comment**, various adjectives that show the subject's appreciation for the application, the system or the experiment; and **Improvement Tips**, which is a list of things that could be improved to make the applications better.

The commentaries related to session comparison (see Table 1 - 1) and the performance of the subject written during the session are ignored, as the subjects did not expect sham accuracy. We have translated the comments from French to English as faithfully as possible. Finally, we added context to comments or shortened them, if necessary, when indicated in brackets.

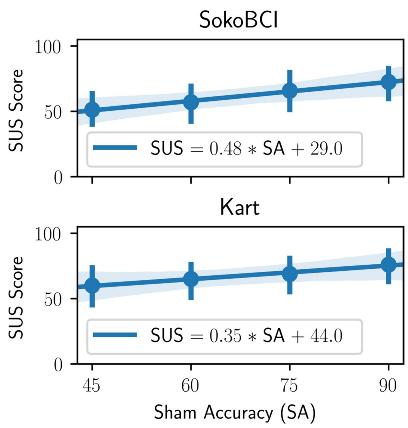


Figure 6: Mean and 95% Confidence Interval, computed by bootstrap, of the SUS Score in regard to sham

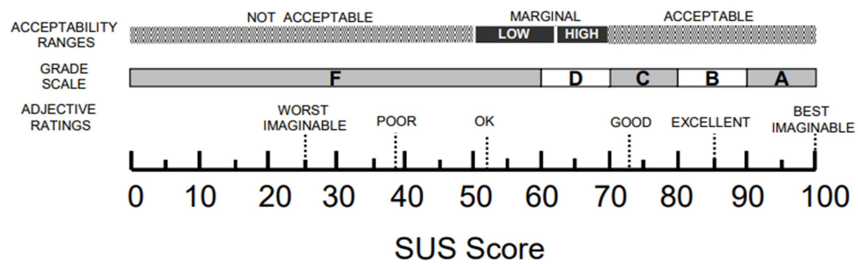


Figure 7: Relationship between the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score. From (Bangor, et al., 2009)

Strategy

Two subjects spontaneously used the same strategy of Motor Imagery during the SokoBCI. They imagined themselves pulling the antenna of the avatar with one or two arms. For example, subject n°10 writes “The design of the character helped me to visualise the correct movements, I imagined taking the antennas on the head of the character and pulling them (the antenna on the right with the right hand to turn to the right, [etc])”.

The subject n°6 comments that he found it easier to focus on the road instead of the kart: “I found it easier to concentrate when I wasn't looking at the vehicle (looking at the road instead) [After Kart 45%].”.

Application Comparison

Subjects 2 and 10 comment about the higher complexity of the SokoBCI over the Kart driving application, subject n°2 writes: “[SokoBCI] is very funny (more than Kart), but more complex too”. However, subject n°4 states during the debriefing having more trouble processing the scene and movement while performing the Motor Imagery task: it's hard to see the game and play at the same time. Acquire information and respond in time, that's difficult.”.

Positive comments

Almost every subject commented something positive; the words “fun” or “playful” appeared a lot. The word “fun” was given by six different subjects. Subject n°10 noticed how time seems to pass much faster and subject n°3 highlights the intuitiveness of the system: “It’s fun finally seeing the system working for me. It’s pretty instinctive to use at this point [after SokoBCI at 75%]”. Subject n°8 comments that the experiment was more ecological: “Last session with a concrete application is a bit more playful.”.

Improvement Tips

Subjects numbers 4, 5 and 8 commented that the inertia of the kart was disturbing them: subject 4 writes "I struggle focusing simultaneously on the stimulations and on the commands to be performed", and subject n°5 writes: "I struggle to focus simultaneously on the stimulations and on the commands to be performed [during Kart]: really tiring. [In SokoBCI] The character is static during the acquisition [i.e. Green light] so we do not receive information, this helps to focus.". Interestingly, subject n°7 mentioned that the breathing animation of the avatar in the SokoBCI during the green light was a distraction: "The [IDLE] movement of the avatar are distracting.".

Subjects 2 and 3 retain some ambiguity on the action to be performed during the Kart as they commented on the questionnaire: "Sometimes, I did not know which movement to imagine. For example: a turn arrives, should I turn by anticipation?" (Subject n°2). In this case, the instruction was repeated, however, the anticipation aspect seems uneasy.

Subject n°4 advised to reduce the duration of the green light periods as it was too long, and subject n°5 advocates for a blue sky in the kart application and wished for a more comfortable EEG headset installation.

CONCLUSIONS, LIMITATIONS AND FUTURE WORKS

In this paper we presented the design of two applications controlled by a BCI: a kart driving application and a puzzle solving application called SokoBCI. The BCI provides four commands. The applications differs on two main points, SokoBCI is more punitive than the Kart in case of mistakes. Indeed, after a mistake the user must react to provide a corrective command in the SokoBCI, while in the kart, invisible walls on each side of the road push back the kart in the good direction as it gets closer in case of mistakes. Secondly, the Kart's motion is continuous and partly controlled by inertia while in SokoBCI the avatar steps from its current position to the next one. It means that the kart keeps moving during the Motor Imagery task and the breaks, and therefore that the user must maintain a significant level of mental activity in the decision-making process to choose the correct command to use.

We evaluated both applications using questionnaires, open written comment and discussions with a group of ten healthy subjects. We observed that some dimensions of the Raw TLX are highly correlated to the performance of the system, like Performance and Frustration, while others don't. In addition, the Mental Demand at high performance of the system, was more important in the Kart than in SokoBCI. It might be explained by the difference in inertia between the application. This was also confirmed by the users in the commentary and discussion.

Additionally, we proposed a model of the relationship between the performance of the BCI and the perceived usability, assessed with an SUS. Firstly, the models show a positive correlation between the variables for both applications. In addition, the correlation was stronger in the SokoBCI than in the Kart. This could be explained by the fact that the SokoBCI is more punitive to the user in case of mistakes. Therefore, one might be concerned by the design choice in the application creation process as it can make an application felt acceptable by the user for a performance level much lower or higher. In our case, we found a ten-point difference in accuracy between the application to reach the acceptability threshold.

One straightforward limitation of this work is the number of subjects included in the experiment, further inclusions are needed to increase the significance of results. Finally, elaborating a more general approach based on characteristics of various application, like inertia or “punitiveness”, to model the impact of those application characteristics on the user might be an interesting tool to help in the design of serious games.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available upon reasonable request from the authors.

ACKNOWLEDGMENTS

The authors would like to acknowledge Camille Bordeau and Philémon Berne, both interns of the team at the time. CB helped the authors by contributing to the conception of the questionnaires and PB assisted the first author in recording from a few sessions of the experiment and greatly contributes to digitising the data from the questionnaires. The authors would also like to thank MD. PhD. Professor Arnaud Delval from the clinical neurophysiology department of the Lille Hospital for his full supports by allowing us to perform the experiments within these infrastructures and with their electroencephalogram system.

RÉFÉRENCES

Bangor, A., Kortum, P. & Miller, J., 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies,* 4(3), p. 10.

Botrel, L., Mira Holz, E. & Kübler, A., 2017. *Using Brain Painting at Home for 5 Years: Stability of the P300 During Prolonged BCI Usage by Two End-Users with ALS.* s.l., Springer International Publishing, p. 292.

Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry,* 189(194).

Brooke, J., 2013. SUS: A Retrospective. *Journal of Usability Studies,* 8(2), pp. 29-40.

Hart, S. G., 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 50(9), pp. 904-908.

Park, S., Ha, J., Kim, D.-H. & Kim, L., 2021. Improving Motor Imagery-Based Brain-Computer Interface Performance Based on Sensory Stimulation Training: An Approach Focused on Poorly Performing Users. *Frontiers in Neuroscience,* Volume 15.