Using a P300 Brain Computer Interface in an Immersive Virtual Environment

Michael Donnerer
m.donnerer@cs.ucl.ac.uk

Anthony Steed
a.steed@cs.ucl.ac.uk

Department of Computer Science, University College London

Abstract

Brain-computer interfaces (BCIs) provide a novel form of human-computer interaction. The purpose of these systems is to aid disabled people by affording them the possibility of communication and environment control. In this study we present experiments using a P300 based BCI in a fully immersive virtual environment (IVE). P300 BCIs depend on presenting several stimuli to the user. We propose two ways of embedding the stimuli in the virtual environment: one that uses 3D objects as targets, and a second that uses a virtual overlay. Both ways have been shown to work effectively with no significant difference in selection accuracy. The results suggest that P300 BCIs can be used successfully in a 3D environment, and this suggests some novel ways of using BCIs in real world environments.

1 Introduction

The definition of a brain-computer interface (BCI) found in Wolpaw et al. (2002) is "A system for controlling a device e.g. computer, wheelchair or a neuroprothesis by human intention which does not depend on the brain's normal output pathways of peripheral nerves and muscles". The term was first introduced in Vidal (1973) where an interface based on electroencephalographic (EEG) signals was proposed.

The main use of BCIs is to provide a means of communication for people who are otherwise unable to do so. One example of this would be to enable patients with locked-in syndrome, whereby nearly all voluntary muscles are paralysed, to communicate with the outside world again. At the moment patients suffering from locked-in syndrome are often only able to communicate through eye movements or blinks. Such people could use a BCI to efficiently use a computer or other devices. A future, though far off, goal of BCI research is to create a communication channel between human and computer that is much faster and more intuitive to use than any current technology.

It is possible to split EEG-based BCIs into four categories based on the EEG features they use: induced changes of oscillatory activity (for examplePfurtscheller and Neuper (2001)), slow cortical potentials (for exampleBirbaumer et al. (1999)), steady-state evoked potentials (SSVEP) and P300 evoked potentials. Induced changes of oscillatory activity and slow cortical potentials are most commonly limited to binary or quaternary choice.

As the name suggests P300-based BCIs are based on the P300 brainwave, which was first discovered by Sutton et al. (1965). It can be evoked by either a visual or auditory stimulus that a user has to concentrate upon while different non-target stimuli are also presented. In Picton (1992) the characteristics of the P300 signal are described more closely. Generally it is generated when an occasional target stimulus is detected by the user among several non-target stimuli. This is called an "oddball" paradigm. To elicit the stimulus the subject has to be actively engaged and the amplitude is larger when the stimulus is less probable.

We are interested in examining the use of a BCI within immersive virtual environments (IVEs). In particular we wanted to integrate the P300 stimuli directly into the 3D environment. If a BCI could be demonstrated in such a way, it would not only have potential benefits for interaction research in IVEs, but also for the development of next-generation BCIs that exploit the real world around the user. For example, an interesting application scenario for a BCI would be to allow the user to control his real world environment. It could be used to navigate the environment in a wheelchair, to select a book in a bookshelf, to turn on the TV and change the channel, et cetera. This could be achieved using an augmented reality style interface, either by interacting with a virtual environment reconstruction, or as a live overlay. Studying BCIs in IVEs provides a test of the feasibility of such techniques.

There are many questions about the feasibility and appropriate types of integration of a P300-based BCI in virtual environments. In this paper we discuss a series of feasibility pilot trials and a main trial that led to the establishment of two paradigms for using a P300-based BCI in virtual environments - the Object Flash paradigm and the Tiles paradigm, which utilizes a virtual overlay. In addition to that, a third paradigm called "36 Spheres" is used for calibrating the system and comparing it to other types of P300-based BCI.

In Section 2 we discuss related work on P300-based BCIs and virtual environments. Section 3 discusses the general setup of our integration of the BCI in an IVE, and the paradigms that we evaluated. Section 4 then discusses pilot and feasibility trials, and Section 5 a more substantial trial. Section 6 then concludes and discusses the potential for future work.

2 Related Work

2.1 P300-based BCI

Farwell and Donchin (1988) presented the first usable brain computer interface based on the P300 signal. They created a 6x6 matrix which was populated with letters and numbers. The rows and columns were then flashed and the users asked to concentrate on a matrix element and count the number of flashes. Thus a P300 signal was generated whenever a row/column was flashed that contained the element the user was concentrating on. Through this, the user was able to spell arbitrary words and sentences. Both online and offline signal detection were tried. In the online condition a rate of 2.3 characters/min was achieved in an experiment with 4 healthy subjects.

There are many influences on the efficiency, i.e. input rate achievable, on P300-based BCI paradigms. The difficulty of discriminating the target stimulus from the background stimuli influences the amplitude of the P300. According to Squires et al. (1973) if the difficulty is too high this leads to a decrease in user confidence that the distinction is correct which leads to a smaller amplitude of the P300 wave. At the same time it has been shown in Hillyard et al. (1971) that a too easy task can also decrease the amplitude. This might be explained by the fact that the mind may stray to other matters if it gets bored. According to Kutas et al. (1977) the amplitude of the P300 wave did not change with the semantic complexity associated with a task but it did influence the latency of the wave. These results suggest that 3D stimuli may need to be carefully designed.

At the moment the exact mechanisms involved in the generation of the P300 wave are still unknown and its precise purpose is still unclear. It is known that it is influenced by alcoholism, schizophrenia, autism and other conditions. Further work on all the criteria influencing P300 elicitation is still needed and this work tries to address some of them.

Hoffmann et al. (2008) did a study using a six choice P300 system using five severely disabled subjects and four able-bodied subjects. The effects of using Bayesian Linear Discriminant Analysis (BLDA) and Fisher's Linear Discriminant Analysis (FLDA) for signal processing as well as the effects of varying the number of EEG electrodes used were investigated. Four of the disabled subjects and all of the able-bodied subjects achieved communication rates and classification accuracies superior to Piccione et al. (2006) and Sellers and Donchin (2006). It was found that using more than 8 electrodes resulted only in a small positive increase in BLDA performance and in a decrease in FLDA performance. It was also noted that muscle activity can cause large amounts of artifacts in the EEG recordings.

2.2 P300-based BCI within Virtual Environments

So far there has been little research in the area of using P300-based BCIs in virtual environments (VEs) and even fewer studies have looked at IVEs. As far as we are aware no previous study has investigated the use of a P300 BCI in a CAVE™-like environment.

Bayliss and Ballard (1999) present an experiment in virtual reality using head-mounted displays. In their experiment subjects were situated in a mock-up of a go-kart and asked to drive through a virtual environment while stopping at red stoplights. The hypothesis was that the red stoplights, being less frequent than yellow and green stoplights, would excite a P300 wave. The EEG data of the subjects was captured and analysed using correlation, independent component analysis, a Kalman filter and a robust Kalman filter. The red lights did excite a P300 wave and the robust Kalman filter gave the best results. One should note that this study used offline analysis only, thus the presence of a P300 was evaluated but it was not used for control in the environment. This aspect was added in Bayliss (2003) where subjects were able to control a virtual apartment. To achieve this, semi-transparent spheres appeared for a short time in front of five controllable items to provide a stimulus.

Besides using a normal computer screen the scenario was also evaluated using a head mounted display. No significant difference was found between the two conditions. This study showed that it is indeed possible to use a P300-based BCI for control in a virtual environment.

The only other study we are aware of using a P300 BCI within a virtual environment is Piccione et al. (2008). However, the study did not utilise a CAVE™-like environment but rather a "3D-view projection display". Details of the display are unfortunately not given in the study so it is not clear if this is an immersive display. In the study four arrows were added to a virtual environment. By counting the number of times a particular arrow flashed the user was able to navigate in the direction the arrow pointed. The study evaluated different arrow models varying in colour and placement and different flash times (70ms and 150ms). They found no significant differences between these conditions. They also compared the virtual environment to a simpler 2D representation, which contained four 2D arrows. They compared the difference between the virtual environment conditions and the 2D condition by computing an "ra" index. This index represented the sum of absolute differences between target and non-target brainwaves recorded. The ra index of the 2D condition is significantly higher than the ra indexes of the virtual environment conditions. Unfortunately, it is not noted if this difference had an impact on selection accuracy. Furthermore, the study showed that no session-by-session improvement in performance occurred, suggesting that user performance cannot be improved by learning or training.

A hybrid approach was presented in Edlinger et al. (2009). In it users were able to control a virtual apartment presented on a stereoscopic Powerwall. This system was not controlled through stimuli in the virtual apartment though. Instead a control matrix was used, which was displayed on a separate screen in 2D. The system utilized multiple control masks, thus allowing many control elements (navigation to rooms, music, windows, heating system, television, etc.). The study is interesting as it presents a practical approach for controlling real world environments by outfitting the user with a separate display to be used as a "remote".

There also have been studies integrating BCIs that are not based on the P300 brainwave with virtual environments. One of these is Lalor (2005) which used steady state visual evoked potential for binary control in a visually elaborate 3D game. The game was presented to the subject on a large screen that did not use stereoscopic display technologies. The goal was to keep a virtual character in balance by either looking at a left or right checkerboard control. It was found that the performance of the BCI was fairly robust against the distracting visual stimulation created by the game.

Similarly Lotte et al. (2008) presented a study "out of the lab". In it users were asked to "use the force" (i.e. a motor imagery based BCI) to lift a model of a spaceship in a 3D environment (non-immersive, non-stereoscopic). 21 naive subjects who were untrained in the use of motor imagery were used. The experiment used only a single EEG channel and showed that some subjects were able to effectively use the BCI with no training.

Furthermore, Leeb et al. (2006) showed that it is possible to move through a virtual environment in a CAVE™-like system through the use of a motor imagery based BCI. Touyama et al. (2008) evaluated steady-state visual evoked potentials in a CABIN environment, which consists of a virtual reality created through the use of five screens. In the experiment the users could choose between two virtual buttons place to the right and left of the user, flickering at different frequencies. The system could classify the two buttons with an accuracy rate of about 70-80%.

3 P300-Based BCI Integration

3.1 System Integration

The overall system setup can be seen in figure 1. The particular IVE system being used was a ReaCTor (SEOS Ltd., West Sussex, UK) system. It consists of three back-projected 3m by 2.2m walls to the left, right and front of the user together with a 3m by 3m front-projected screen on the floor. Stereo shutter glasses are used to separate the left and right eye images, which are displayed at 45Hz each. The user's head position is tracked through an Intersense IS-900[1] system, which is connected to an SGI Onyx2, which for these trials only serves tracking data. The tracking information is then passed on to the ClusterMaster computer, which is equipped with 2GB Ram and dual 1.8GHz Intel processors running Windows XP. ClusterMaster executes our software for displaying the virtual environments. The software was written using VRMedia XVR[2]. The XVR application on ClusterMaster delegates the rendering to the cluster slaves. Each cluster slave is equipped with 1GB RAM, a single 2.7GHz Intel processor and a GeForce Quadro 5600 graphics card and runs Windows XP.

During the experiments the users were seated on a chair located about 60cm from the back side of the IVE equidistant to either side wall. A g.Mobilab+ device transmitted the EEG-data via Bluetooth to the portable computer positioned behind the user, which was running the signal processing and control code. This in turn sent instructions to the XVR program on ClusterMaster via a local network connection.

[1] http://www.isense.com/

[2] http://www.vrmedia.it

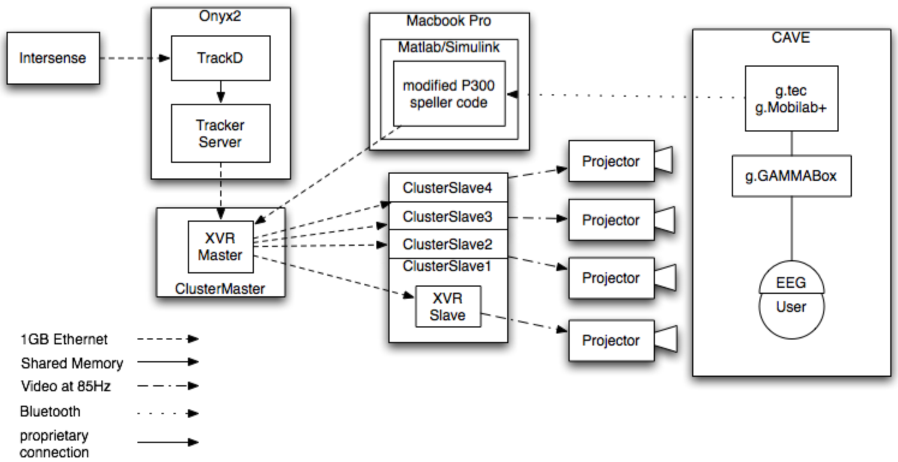

Figure 1: Systems diagram for our CAVE™-like system

Signal processing was handled by a modified g.tec P300 speller[3] using linear discriminant analysis. The hardware used consisted of a g.tec g.Mobilab+ EEG device with active electrodes powered by a g.GAMMAbox. Eight electrodes with one reference and one ground were used. The electrodes were mounted on Fz, Cz, P3, Pz, P4, PO7, Oz and PO8 (according to the extended 10-20 system presented in Jasper (1958)) with reference mounted on the right ear lobe and ground mounted on the forehead (AFz).

3.2 P300-Based BCI Paradigms

Three different paradigms were investigated: *36 Spheres*, *Object Flash* and *Tiles*. They can be seen in figures 2, 3 and 4, respectively. 36 Spheres is a straightforward "translation" of the Farwell and Donchin (1988) speller into a virtual environment. The main difference between Farwell and Donchin (1988) and 36 spheres and our other paradigms is that we only flash a single object at a time instead of a whole row/column. We chose this because we thought it would make for an easier translation into a virtual environment, where it can be difficult to identify obvious choices for mapping rows/columns. A solution based on multiple object might promise higher selection throughput and would thus be an interesting

[3] http://www.gtec.at/service/Tutorials/P300SpellerwithgUSBamp.pdf

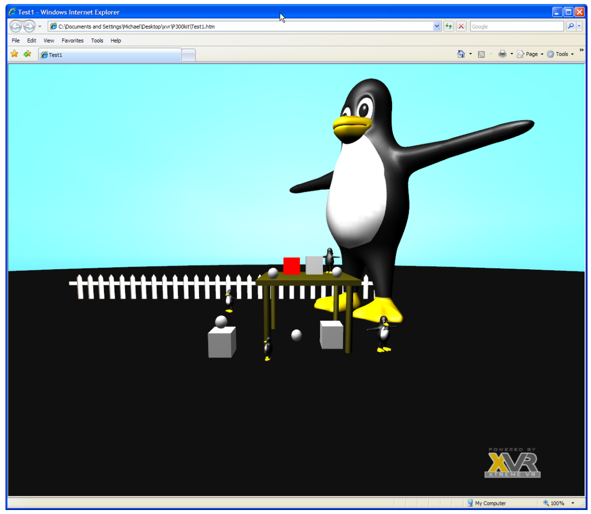

Figure 2: Object Flash Scene - a cube is currently being flashed

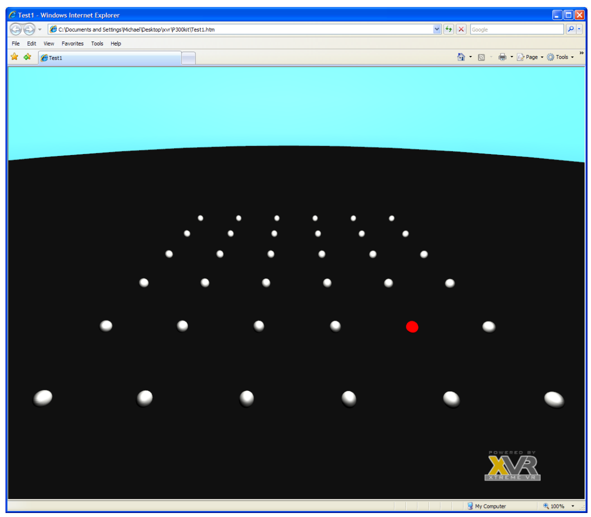

Figure 3: 36 Spheres Scene - a sphere is currently being flashed

target for future research. In the 36 spheres paradigm, 36 spheres are arranged in a regular grid. Each sphere can flash in colour. The closest spheres were about 3.5 m away from the user with a side of the rectangle formed by all spheres measuring approximately 2m. The spheres are slanted by 45 degrees, such that rows further away are higher. We only used 36 Spheres for the generation of a classifier for the other two scenes. This choice was made as 36 Spheres proved to work quite well for this purpose in our pre-trial. To generate a classifier the participants were asked to count the number of flashes on each sphere in the bottom row in turn - 16 flashes per sphere were used.

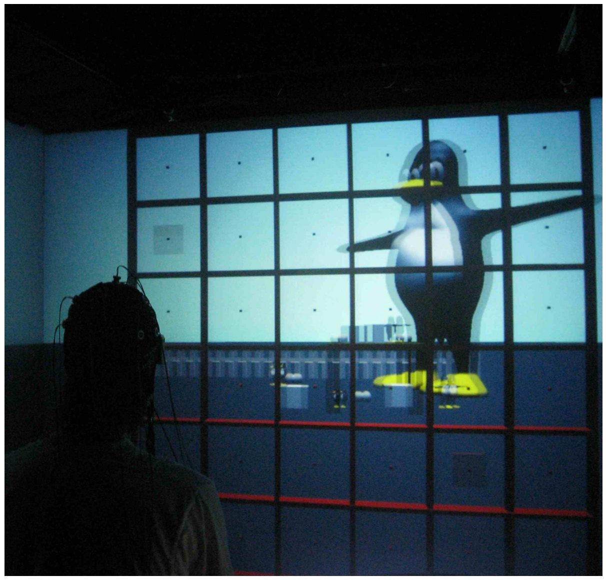

Figure 4: Tiles Scene presented to a subject in the IVE. Due to the long shutter time of the camera multiple left and right eye images are seen in the picture and thus two tiles are visible at the same time.

Object Flash takes 36 Spheres one step further. Instead of having uniform objects arranged in a regular grid, cubes, spheres, penguins, a table and a fence are arranged in an irregular way. The scene features some highly visually cluttered areas, especially around the table-top. The closest object is about 6.5m away from the user's head, with the table approximately a meter high (life size). The stimuli are provided by flashing the individual objects in red. In figure 2 one of the cubes on the table is currently being flashed.

Tiles uses the same basic scene as Object Flash. The difference is the way the P300 stimulus is elicited. Instead of flashing the individual objects, the user looks onto the scene through a virtual overlay appearing as a segmented window. The window is segmented into 36 different areas, in each area a "tile" can appear randomly. The P300 brain wave is thus elicited by having the user concentrate on the appearance of a tile in a particular segment. Each segment also provides a focus point at the center consisting of a small cube. Furthermore, the segments are separated by a visible grid. This setup is a direct result of our pre-trial, which also evaluated different possibilities (see section 4). The grid is positioned about 1.5m away from the user, lying roughly in the plane of the front screen, being about

2x2m in size.

As mentioned previously a classifier was generated using 16 flashes, all other conditions used 8 flashes per object. One object was flashed at a time. The scene presented to the user had hidden objects that were also flashed, thus yielding 36 possible objects in every scene. These hidden objects could be selected unintentionally. Their main purpose was to make the timing between the different paradigms comparable. Also they afford us the possibility to investigate the difference between random errors and errors caused by the user inadvertently reacting to a non-target stimulus. The flashes in all scenes were pure red, although in the Tiles paradigm they appeared dark red due to the presence of shadows from the grid. After signal processing, selection results were displayed in two different ways. In Object Flash the selected object was flashed for two seconds using a green flash. In Tiles the selected tile was displayed for two seconds using a green tile.

The flash-time was set to 45ms in all conditions, with the actual time varying slightly depending on the current frame rate. We measured the frame rate during a trial run. On both Object Flash and Tiles the framerate stabilised at around 20 FPS per eye.

Each flash is delivered 70 ms after the previous one, giving a “dark time” of 25ms. Of course, the actual display timings vary due to the jitter resulting from the network connections and the fact that a frame rate of 20 FPS affords only a timer granularity of 5ms. Once all the flashes necessary for one selection have been displayed, the system processes the data and provides a result in about a second. There is a break of 3 seconds after all the flashes for one round have been shown. During this break the selection result is displayed for 2 seconds. Using 8 flashes per object, each selection takes 20 seconds. Adding the time used for displaying the result, one selection takes 23 seconds.

3.3 Jitter and Latency

The complicated system setup involving multiple computers and networks as well as a relatively low frame rate of 20 FPS in the experiment introduced additional jitter and latency.

As it is crucial for correct system operation to be able to correlate the time a stimulus was displayed with the time a P300 brainwave was detected, jitter and latency can negatively impact performance. In our pre-trial a single subject test suggested that a latency of up to 100ms and jitter of up to 20ms can be tolerated by the signal processing code without significantly impacting performance. The test was performed by passing the network connection through a proxy which could delay the packets for an adjustable amount of latency and jitter.

To further combat latency we modified the signal processing code to account for a variable amount of latency. In our experiments this was set to 40ms. This value was chosen due to two reasons. First, the end-to-end latency of the system was measured to be 64ms using the method presented in Steed (2008). The method cannot be used directly to estimate the latency of the whole bio-processing system as we have to discount the latency of the tracker and add the latency introduced by the additional network connections. Second, a value of 40ms proved to be effective in our pre-trial. Therefore, as our latency tests indicated that a small amount of latency is fine, we decided to run the system at 40ms latency correction. Assuming that the actual latency is in the area of 40-80ms, 40ms is a good choice as it will not correct more latency than actually exists, while still reducing the actual latency considerably.

4 Pilot Trials

Before the main experiment a series of pilot trials were run to find good settings for the stimuli for the 36 Spheres, Object Flash and Tiles paradigms. These pilot trials used both immersive and non-immersive settings. There are several factors that could bias the results, and we sought to identify any obvious confounding factors. For example, the size, distance and colour of the objects could all potentially affect the efficiency of the BCI. If the efficiency would be too greatly affected, then the aim of integrating P300-based BCI in to a real-world

subject	white	red	mean
1	6	6	6
2	5	6	5.5
3	6	6	6
mean	5.7	6	5.8

Table 1: Number of spheres correctly selected in 36 Spheres with 6 tries in the non-immersive setting

setting where such factors can not be controlled, is more likely to be unachievable.

Three subjects were used in the non-immersive setting and three in the immersive setting. Each experiment lasted for approximately 50 minutes and subjects were not compensated for their time. Non-immersive participants did all three paradigms whereas immersive participants did only the 36 Spheres and Object Flash paradigms. The reason for this is that at the time of the pilot trials we had not created the immersive tiles condition but instead tested other conditions which we thought to be superior. The results for these conditions were not very good so we decided to implement an immersive tiles condition instead. In addition to that, both groups of subjects did further tests, which do not relate to the paradigms presented in this paper.

In the 36 Spheres we tested the difference between a white and a red flash. As the spheres themselves are white, the white flash is much lower in contrast than the red flash. We wanted to investigate whether this difference in contrast would influence selection accuracy. The results showed only a small difference which can be seen in table 1. One subject expressed a preference for the white flash, describing the red flash as "mesmerizing". The two other subjects indicated that they preferred the red flash as it was easier to detect.

In the immersive condition different sizes for the spheres and different distances were evaluated. The results can be seen in table 2. The results indicate that large spheres (about 30cm in diameter) close to the user (about 50cm away) provide lower accuracy rates. Due to this we did not use this close condition in the main trial.

The results for Object Flash in the non-immersive condition can be seen in table 3. Table 4, 5 and 6 show the results for Object Flash in the immersive scene for a close (about 80cm

Subject	far	close-big	small	mean
4	2	1	4	2.3
5	3	0	6	3
6	6	6	6	6
mean	3.7	2.3	5.3	3.8

Table 2: Results for selecting 6 spheres in an immersive environment

subject	penguin	table	cube	mean
1	3	1	3	2.3
2	3	2	2	2.3
3	3	1	2	2
mean	3	1.3	2.3	2.2

Table 3: Number of objects successfully selected in the Object Flash paradigm with 3 trials in the non-immersive condition

from the user), medium (about 3m) and far condition (about 6.5m).

Finally, table 7 shows the results for the non-immersive version of Tiles, which has the same functionality as the immersive version but a slightly different look. A mean accuracy of 2.7 out of 3 selections was achieved, which is higher than the mean accuracy for Object Flash (0.9).

5 Main Trial

The main hypotheses for the study are that both the Object Flash and Tiles paradigm can be used successfully in an IVE with the Object Flash paradigm providing lower accuracy rates for highly cluttered areas. At the same time Tiles should be immune to visually cluttered environments and provide similar or higher performance as Object Flash.

subject	penguin	table	cube	mean
4	1	0	0	0.3
5	1	0	0	0.3
6	2	1	3	2
mean	1.3	0.3	1	0.9

Table 4: Results for Object Flash Immersive under the close condition with 3 trials

subject	penguin	table	cube	mean
4	0	0	0	0
5	3	0	1	1.3
6	3	0	0	1
mean	2	0	0.3	0.8

Table 5: Results for Object Flash Immersive under the medium condition with 3 trials

subject	penguin	table	cube	mean
4	0	0	0	0
5	3	0	1	1.3
6	3	0	1	1.3
mean	2	0	0.6	0.9

Table 6: Results for Object Flash Immersive under the far condition with 3 trials

subject	tile 1	tile 2	tile 3	mean
1	3	3	3	3
2	1	2	3	2
3	3	3	3	3
mean	2.3	2.7	3	2.7

Table 7: Number of cells correctly selected in the Tiles experiment with 3 trials in the non-immersive setting

5.1 Subjects and Methods

The main trial involved seven participants in the age range of 20-25 with two female and five male subjects. Each experiment took 40 minutes and subjects were not compensated. Each participant was first asked to generate a classifier by selecting the bottom six spheres in the 36 Spheres condition. It was chosen to use the bottom row of spheres as it is easy for users to select them as they don't have to search for the next sphere as it will be the one immediately to the right of the current sphere. In addition to that the bottom row was closest to the user making the perceived size of the spheres bigger than that of the spheres in the other rows, helping participants more easily see a flash. We decided against creating multiple classifiers for the different paradigms even though this might have improved results. Our main worry was that the lower and more varied frame rate in the other paradigms could impact the P300 classifier generation negatively.

Participants were tested on the Object Flash and Tiles paradigm in randomized order. The positions of all the objects and the tiles were kept the same for all participants. In the Object Flash condition the participants were asked to select three objects - the large penguin, the table and the cube to the left of the table five times each. Each object was selected five times before the next object was selected. The order the objects were asked to be selected in was randomized. In the Tiles paradigm the participants were asked to select 15 tiles, starting at the second row from the top at the left, moving one to the right after each selection (no matter whether the selection was successful or not), wrapping around to the next row when the end of a row was reached. One of the tiles was very difficult to see as it was directly in front of a very dark area of the large penguin. All other tiles were clearly visible.

After the experiment the participants were asked to fill out a simple questionnaire. The first question asked how easy it was to concentrate on a particular object (large penguin, table, box, tiles) regardless of the result achieved. It was possible to choose between the following answers: very easy, easy, neutral, difficult, and very difficult. The second questions

subject	table	penguin	box	mean
1	1	5	5	3.67
2	0	1	1	0.67
3	5	5	5	5
4	0	1	4	1.33
5	0	2	2	1.33
6	3	2	0	1.67
7	2	5	4	3.67
mean	1.57	3.00	3.00	2.48
variance	3.62	3.67	4	

Table 8: Objects successfully selected in the Object Flash paradigm using 5 tries

asked how much the user enjoyed selecting the penguin, table, box and tiles and provided the following answer possibilities: a lot, a bit, neutral, not really, not at all.

5.2 Results

Table 8 shows the results for selecting the different objects in the Object Flash scene. The differences in selection success rates between the table, penguin and box are not significant ($p<0.05$). Table 9 shows the results for the Tiles paradigm compared with the overall results of the Object Flash paradigm. Again the difference is not significant. The results from the questionnaire are in table 10 and 11. The five different answer possibilities were assigned numeric values from +2 to -2. Table 10 shows the results for the perceived easiness. Using a 3-way ANOVA on table, penguin and box with a Bonferroni correction and $p<0.05$ the difference between the table and the penguin is significant with the penguin being "easier" to select. Without the Bonferroni correction the penguin was also significantly "easier" to select than the box. Table 11 shows participant enjoyment the different conditions. Using a 3-way ANOVA on table, penguin and box with a Bonferroni correction and $p<0.05$ the difference between table and box is significant with the box being more "enjoyable". Without the Bonferroni correction the difference between penguin and table is significant as well with the penguin being more "enjoyable".

Table 12 presents the results for selecting the first tile and the tile which was hard to

subject	Object Flash	Tiles
1	11	13
2	2	2
3	15	13
4	5	11
5	4	11
6	5	6
7	11	13
mean	7.57	9.86
variance	22.62	18.14

Table 9: Results of Object Flash compared with Tiles using 15 tries

subject	large penguin	table	cube	tiles
1	2	0	1	1
2	2	-2	1	2
3	2	1	1	0
4	1	1	2	0
5	2	-2	-1	2
6	1	0	0	1
7	2	-2	1	2
mean	1.71	-.57	.71	1.14
variance	0.24	1.95	0.9	0.81

Table 10: Results of the questionnaire on how easy was it to concentrate on the particular objects

subject	large penguin	table	cube	tiles
1	2	-1	1	1
2	1	-2	1	1
3	2	0	0	-1
4	-2	0	2	0
5	2	-2	2	2
6	2	-1	-1	-1
7	0	-2	1	2
mean	1	-1.14	.86	.57
variance	2.33	0.81	1.14	1.62

Table 11: Results of the questionnaire on participant enjoyment of the conditions

subject	first tile	"difficult" tile
1	1	0
2	0	0
3	1	0
4	1	1
5	1	0
6	0	0
7	1	0
mean	.71	.14
variance	.24	.14

Table 12: Comparison of the selection rate for the first tile in the 2nd row and the "difficult" tile which was hard to see

see. The difference between the two is significant (t-test, p=0.03) with the hard to see tile providing less successful selections.

For the tiles condition we also did an analysis of the mis-classifications. 11 of the 38 mis-classifications are on a neighbouring letter and 27 aren't. On average each tile in the 15 selected tiles has 7 neighbours. Thus 28.94% of the mis-classifications were a neighbour which make up only 19.44% of the possible false tiles.

Table 13 and 14 compare our results with the best results from Bayliss (2003). In Bayliss (2003) it was only possible to select between five objects. Both our systems allow a selection of one from 36 even though the Object Flash paradigm only visibly shows 15 objects. Because of this our system achieves lower selection rates per minute as there are more objects to be flashed. We calculated bit rates for our system by taking the log_2 of the number of possible targets and multiplying this number by the average selection accuracy. The bit rates achieved are also compared with the online bit rate achieved in Serby et al. (2005) which uses a Farwell and Donchin (1988) type 2D speller. Looking at the mean bit rate achieved Object Flash fares worse than the best result from Bayliss (2003) while Tiles achieves a higher bit rate. All results are lower than the ones achieved in Serby et al. (2005). We think that the bit-rate could be improved if more flashes per object were used as the percentage of objects correctly selected was quite low for many participants.

We also analysed the errors in Object Flash by splitting them into two categories: se-

	Bayliss	Object Flash	Tiles
number of goals/minute	3.16	1.21	1.58
variance	1.15	.76	.68

Table 13: Number of objects successfully selected per minute compared with the best result from Bayliss (2003)

	Bayliss	Object Flash	Tiles	Serby et al.
bit rate/minute	7.34	4.73	8.17	15.3

Table 14: Mean bit rate calculated using the number of objects successfully selected per minute compared with the online results achieved in Bayliss (2003) and Serby et al. (2005) which used a Farwell and Donchin (1988) type speller

lecting another visible object and selecting an object that is not in the scene (as only 15 objects are in the scene while the signal processing still runs with 36 objects - the rest being "hidden objects"). In this case 28/53 were selecting another visible object. Thus 52.83% of errors were made on 41.67% of the objects suggesting that there is an error bias caused by the flashing of other objects in the scene.

6 Discussion and Conclusion

Both the paradigms we presented, Object Flash and Tiles, have been shown to work in a fully immersive environment.

Although the differences between the selection rates for different objects were not significant in the Object Flash paradigm, we strongly believe that objects in highly cluttered areas are harder to select. This is based on three observations. First, our pre-trial actually produced a significant difference between the large penguin and the table. Second, Cinel et al. (2004) supports the assumption that near-target stimuli worsen results as they trigger a wave similar to the P300 wave. Third, in the analysis of errors in the Tiles paradigm, neighbouring cells were incorrectly selected more often than expected if errors were assumed to be randomly distributed. In addition to that 5 of the 7 subjects achieved lower selection accuracies on the table than on either the penguin or the box, one subject selected all three objects completely successfully and only one subject scored better on the table. Thus it seems quite likely that flashes close to the target flash distract the user and lead to higher error rates. Based on this it is reasonable to assume that visually cluttered areas in Object Flash will lead to more selection errors.

The size of the object being flashed does not seem to create a real difference in selection accuracy as both the big penguin and the much smaller box achieved similar classification accuracies.

In our system a certain amount of jitter arose because of the use of several network connections to integrate the different systems. Based on the volumes of data transferred, most of the jitter is likely coming from the PC cluster-based rendering system. Unfortunately the renderer system of the XVR package was not open source preventing us from investigating it, but we suspect there may be some occasional frame-long delays in the cross-display synchronisation.

We think that the biggest problem with the two paradigms in a CAVE™-like system is caused by the latency and jitter inherent in the system. While we were able to correct for latency in a certain way, correcting for jitter is much harder. We think that there might be two possible ways of dealing with jitter. First, it would be possible to have a world clock signal that is delivered to all computers in the system. Based on this world clock signal all parts of the system then guarantee that events will happen in a predictable amount of time. However, this approach necessitates extra hardware and quite possibly real-time operating systems. Second, it would be imaginable to monitor jitter in real-time and feed this back into the P300 analysis software. This could be achieved by filming the CAVE™-like screens and analyzing the jitter and taking the actual display timings into account when doing the signal processing. Another possible factor influencing performance is the low brightness and thus the low contrast of the immersive display. We do think this has been avoided through the use of the high contrast stimuli.

A possible way of further improving the performance of the two paradigms would be to flash multiple objects at the same time instead of only flashing a single object/tile. We think that using this, it might be possible to achieve higher accuracy than the corresponding single flash model, while still taking less time. This does not conflict with the result in Guan et al. (2004) as they used shorter duration flashes in the single flash model, which could also be applied to the row/column model.

Comparing the Object Flash paradigm with the Tiles paradigm, it seems like the latter is the better in highly cluttered environments to avoid the presence of too many flashes being too close together and thus negatively affecting performance. Of course, the actual object selection is much simpler in Object Flash than in Tiles. To successfully select an object in Tiles the user selects the tile the object can be seen in. However, there might be multiple objects in the same cell and this also depends on the users head position. Taking the users head position into account can be easily done in CAVE™-like environments through the use of head-trackers which are already used to generate the correct perspective images. Through this it would also be possible for the user to help disambiguate the display by actively moving his head. In addition to that, dealing with multiple objects per cell could be done in two ways. First, a zoom could be performed on the selected tile until it fills the whole screen. Then the user is allowed to select again - this time in a much finer manner. This can be repeated until there is only one object in the tile selected or until one object can be classified as the "dominant" object. A second option would be to display a smaller version of Tiles at the position of the selected tile allowing sub-selection. This smaller version of Tiles could feature fewer cells depending on how many different objects the user can see in the particular cell. This approach is limited though in that the cells might become too small to select successfully.

In summary, both the paradigms presented demonstrate the feasibility of using a P300 BCI in an immersive scene. Whilst our current work has looked at fully immersive virtual environments, we believe these paradigms could be transferred to other 1st person paradigms, such as interaction with the real world through augmented reality displays, or projected augmented reality.

Acknowledgements

Many thanks to Beste Yuksel for helping in the experiments and to Josephine Harris for taking pictures of the experiments.

References

Bayliss, J. (2003). Use of the evoked potential P3 component for control in a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):113–116.

Bayliss, J. and Ballard, D. (1999). Single trial p300 recognition in a virtual environment. *Computational Intelligence: Methods & Applications Proceedings, Soft Computing in Biomedicine.*

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., and Flor, H. (1999). A spelling device for the paralysed. *Nature*, 398(6725):297–298.

Cinel, C., Poli, R., and Citi, L. (2004). Possible sources of perceptual errors in P300-based speller paradigm. *Biomedizinische Technik*, 49:39–40.

Edlinger, G., Holzner, C., Groenegress, C., Guger, C., and Slater, M. (2009). Goal-oriented control with brain-computer interface. *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, pages 732–740.

Farwell, L. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523.

Guan, C., Thulasidas, M., and Wu, J. (2004). High performance P300 speller for brain-computer interface. In *2004 IEEE International Workshop on Biomedical Circuits and Systems*.

Hillyard, S., Squires, K., Bauer, J., and Lindsay, P. (1971). Evoked potential correlates of auditory signal detection. *Science*, 172(3990):1357–1360.

Hoffmann, U., Vesin, J., Ebrahimi, T., and Diserens, K. (2008). An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience methods*, 167(1):115–125.

Jasper, H. (1958). The 10-20 electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375.

Kutas, M., McCarthy, G., and Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 197(4305):792–795.

Lalor, E. (2005). Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment. *EURASIP journal on applied signal processing*, 2005(19):3156–3164.

Leeb, R., Keinrath, C., Friedman, D., Guger, C., Scherer, R., Neuper, C., Garau, M., Antley, A., Steed, A., Slater, M., et al. (2006). Walking by thinking: the brainwaves are crucial, not the muscles! *Presence: Teleoperators and Virtual Environments*, 15(5):500–514.

Lotte, F., Renard, Y., and Lecuyer, A. (2008). Self-Paced Brain-Computer Interaction with Virtual Worlds: A Quantitative and Qualitative Study Out of the Lab. In *Proc. 4th Int'l Brain-Computer Interface Workshop and Training Course*.

Pfurtscheller, G. and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134.

Piccione, F., Giorgi, F., Tonin, P., Priftis, K., Giove, S., Silvoni, S., Palmas, G., and Beverina, F. (2006). P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clinical neurophysiology*, 117(3):531–537.

Piccione, F., Priftis, K., Tonin, P., Vidale, D., Furlan, R., Cavinato, M., Merico, A., and Piron, L. (2008). Task and Stimulation Paradigm Effects in a P300 Brain Computer Interface Exploitable in a Virtual Environment: A Pilot Study. *PsychNology Journal*, 6(1):99–108.

Picton, T. (1992). The P300 Wave of the Human Event-Related Potential. *Journal of clinical neurophysiology*, 9(4):456.

Sellers, E. and Donchin, E. (2006). A P300-based brain–computer interface: initial tests by ALS patients. *Clinical Neurophysiology*, 117(3):538–548.

Serby, H., Yom-Tov, E., and Inbar, G. (2005). An improved p300-based brain-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 13(1):89–98.

Squires, K., Hillyard, S., and Lindsay, P. (1973). Vertex potentials evoked during auditory signal detection- Relation to decision criteria. *Perception and Psychophysics*, 14:265–272.

Steed, A. (2008). A simple method for estimating the latency of interactive, real-time graphics simulations. In *VRST '08: Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 123–129.

Sutton, S., Braren, M., Zubin, J., and John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188.

Touyama, H., Aotsuka, M., and Hirose, M. (2008). A Pilot Study on Virtual Camera Control via Steady-State VEP in Immersing Virtual Environments. In *Proceedings of the Third IASTED International Conference*, volume 611, page 43.

Vidal, J. (1973). Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180.

Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767 – 791.

List of Figures