Evaluation and Comparison of a Multimodal Combination of BCI Paradigms with Consumer-Grade Hardware and Eye Tracking

Nataliya Kos'myna

LIG-IIHM

41 rue des mathématiques 38400 Saint Martin d'Hères Nataliya.Kos-Myna@e.ujfarenoble.fr

Franck Tarpin-Bernard

LIG-IIHM

41 rue des mathématiques 38400 Saint Martin d'Hères franck.tarpin-bernard@imag.fr

Abstract

This article focuses on multimodal combinations of BCI and eye tracking in the context of a simple puzzle game involving tile selection and rotations using consumergrade EEG hardware. We present our experiment performed on 30 Subjects and present the results from this preliminary study. We come to the conclusion that while eye tracking alone remains a more robust modality, the addition of BCIs with inexpensive hardware brings some interesting properties. First of all, the performance of Eye Tracking (ET) + SSVEP is quite close to that of unimodal eye tracking. Furthermore despite lower performance of ET+ Motor Imagery due to the limitations of the hardware, the users appreciated the interaction modality and thought it was the most natural.

Author Keywords

Multimodal interaction; evaluation; Brain Computer Interfaces; Eye Tracking

b) Shifted Placement

Figure 1. 10-20 Electrode

placements for Emotiv Epoc

ACM Classification Keywords

H.5.2. User Interfaces Input devices and strategies

General Terms

Experimental; Measurement

Introduction

Brain computer interfaces (BCI) first appeared in the 1970s and required brain implants. Thanks to Electroencephalography they have become adapted for more general-purpose applications. Besides expensive high medical devices, some quality affordable Electroencephalography (EEG) headsets, such as Emotiv Epoc [4], claim to allow for the use of BCIs in mainstream applications such as gaming. Because they are very easy to use (wireless, no gel), these kinds of devices have the opportunity to fit with the basic requirements of a gaming experience for non-disabled people. However, BCI systems still suffer [11] from many issues (high error rates, insufficient bandwidth, long learning phases) and these problems are exacerbated with consumer-grade hardware.

An interesting approach appears to be the use of BCIs in multimodal interaction systems to achieve either redundancy or complementarity (in terms of CARE properties [2]).

Among all the possible modalities that could be used in conjunction with BCIs, we have chosen to start with another modality that does not involve any physical contact as such: Eye Tracking [7]. The choice of Eye Tracking seemed pertinent, as not only has it been used for over 20 years in Human Computer Interaction (HCI) [3], but it also appears fairly robust. More importantly, eye tracking exhibits properties very complementary to BCIs. Indeed, eye tracking is is

naturally adapted to selection tasks, while BCIs are more adapted for the triggering of actions.

In this context, we performed an experiment in order to evaluate the effects and performance of multimodal combinations of BCIs and eye tracking using affordable consumer-grade EEG equipment.

We first briefly introduce the topic and review related research. Then, we describe the equipment used, followed by the motivations that led to the choices of said equipment, software platform and the interaction modalities considered. Subsequently, we describe our equipment motivations. Before analysing the results both quantitatively and qualitatively, we give a more in-depth view of the interaction modalities (acquisition, signal processing, classification performance). Finally, we conclude on our findings and draw a plan of action for more extensive future experiments.

Background

We will here briefly present some background information that introduces the key ideas that we exploit and expand upon.

Eye tracking

An eye tracker, as the name suggests is a device that allows the tracking of the coordinates of where a user is directing his/her gaze on a computer screen. While there has been a long history of using eye tracking in Human Computer Interaction [3], its use for the interaction with games is recent [13].

Brain Computer Interfaces

We are using two popular interaction methods suitable for rotation tasks: Steady State Visual Evoked Potentials (SSVEP) [9] and Graz motor imagery [10]. SSVEP is based on the evocation of potentials at a certain frequency stimulated by a flickering target at

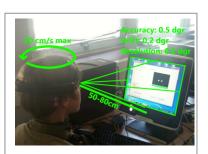


Figure 2. The Tobii T60 Eye Tracker

that same frequency. As for Graz, it is based on imagined hand movements that allow for left and right interactions through direct feedback.

Multimodal BCI Interfaces

As mentioned before, due to the limitations of BCI interfaces [11], researchers have taken interest in combining BCIs with more robust modalities [5][6]. For example, Vilmek and Zander have proposed the combination of eye tracking and BCIs [14] as a spelling system (respectively for selection and validation).

To our knowledge there have been no attempts to combine these two modalities in the context of games strictly speaking, nor have there been an evaluations of different BCI paradigms. However, there have been studies of such combinations for the assessment of executive function [1]: they experimented with a combination of eye tracking and P300 and found that the combination of modalities is beneficial and usable, but that the accuracy of the BCI calibration was a crucial element to the performance of the system.

Equipment

We use the Emotiv Epoc wireless and portable EEG acquisition headset [2] sold by Emotiv Systems.

The Epoc has 14 channels (saline electrodes), as well as two reference channels: AF3, AF4, F3, F4, F7, F8, FC5, FC6, P3, P4, P7, P8, T7, T8, O1, and O2 as illustrated in **Figure 1.a**.

The sampling is sequential at a rate of 128 Hz with an effective resolution of 14 bits. The bandwidth is 0.2 - 45Hz and the dynamic input range is 256 mVpp.

As for the eye tracking, we had at our disposal a Tobii T60 eye tracker (**Figure Figure 2**), which uses infrared VoG (Video-Oculography, eye tracking based on filing the motion of the eyes) to track gaze coordinate.

The T60 offers a data rate of 60Hz, with an accuracy of 0.5 degrees, a drift of 0.1 degrees, a spatial resolution of 0.2 degrees, and a head movement error rate of 0.2 degrees.

The head of the subject is allowed to move with no interruption of the eye tracking within a window of 44cmx22cm at a 70cm distance from the tracker, with a supported distance range between 50cm and 80cm, and a maximum head speed of 20cm/s.

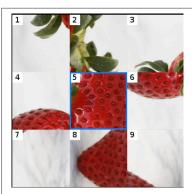
The latency is 33ms, with a blink recovery time 17ms, and a 300ms recovery time when the tracking is interrupted.

Motivations

The objective of these preliminary experiments is to determine the worth of consumer grade EEG headsets for a simple application through the multimodal combination with a more robust interaction modality: eye tracking. Through this approach we hope to obtain an improvement over the performance of BCIs alone and to provide a more natural interaction than Eye tracking alone for certain tasks that do not simply come down to a basic selection.

First, our choice of the Emotiv Epoc headset is mainly motivated by its low cost and by the convenience of the installation procedure compared to medical-grade gel electrodes. Even though the electrode placement on the Epoc headset is hardly suitable for Graz (motor cortex coverage), we considered an alternative headset placement to enable us to still consider Graz for our experiments, as detailed previously. Furthermore, with the default positioning, SSVEP can be used very comfortably.

As for the choice of the eye tracking system, we initially considered using webcam based eye tracking. However, the precision of the tracking was very low and led to selection tasks being practically impossible to perform. Therefore, we settled on using a classical eye tracker. Given that we had at our disposal a Tobii T60 eye tracker, we decided to make use of it for our experiments (See the **Equipment** section).





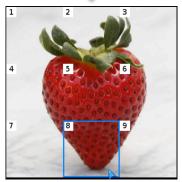


Figure 3. From the start state to the final state of the game

For the BCI modalities, we wanted to evaluate a representative spectrum of the main BCI paradigms and as such we wanted to use Graz, SSVEP and P300. However, due to the fact we were communicating with OpenViBE through the VRPN protocol, the latencies encountered made the use of P300 extremely difficult. Thus, for this first step, we only considered SSVEP and Graz.

Our Test Application

In order to evaluate the various modality combinations and using the software framework of OpenViBE [12], we were able to build a simple test.

The application takes the form of a simple puzzle game, the solving of which constitutes the objective for our test subjects.

The application features an image split into 9 tiles, each rotated in the wrong orientation. The objective of the game is to rotate each tile to put it back in the proper orientation and thus reconstitute the original image, as illustrated by **Figure 3**.

We first implemented a mouse-only version of the game where for testing and instructional purposes. The cursor serves to select a tile; left and right clicks serve to rotate a tile left or right respectively.

Let us now review the different modality combinations that we considered in the application and for the experiments.

Multimodality and Fusion

Fusion Engine and CARE Properties

Before describing the modality combinations, it is important that we describe more formally the nature of the combinations by using some of the existing frameworks relating to multimodal interaction.

In the case of eye tracking combined with SSVEP and of eye tracking combined with Graz, we use and alternative modality fusion scheme. Following the classification criteria presented in [8] we implemented a procedural notation-less fusion, at a low level for eye tracking (coordinates) and at the dialog level for the BCI modalities (left/right actions).

Pure Eye Tracking

The unimodal eye tracker version uses eye tracking both for selecting the puzzle tiles and rotating them. The selection and rotation phases are successive and disjoint. First, the user is instructed to select a tile and is granted 10 seconds to proceed; it then becomes possible to rotate the tile during 10 seconds (by looking to left or to the right of the selected tile), followed by a 5 second rest period.

Eye Tracking + SSVEP

The selection phase remains the same as with the eye tracker only version. For the rotation phase, arrows appear on both sides of the tile and flicker at 15 and 30Hz respectively. By looking at the arrow corresponding to the desired rotation direction, evoked visual potentials are generated and then processed to trigger the rotation command for the game. Each phase was set to last 10 seconds followed by a 5 second rest.

Eye Tracking + Graz Motor Imagery

Individually, the selection phase and rotation phases each last 10 seconds, with a rest period of 10 seconds minimum (we use this to synchronise the application with the Graz protocol in OpenViBE). For the rotation phase, a feedback bar representing the degree of left or right classification is shown to the user. The side on which the feedback value dwells the most on is considered as the selected rotation direction.

Contrarily to SSVEP that cannot be used simultaneously with the eye tracker, it is possible when using Graz. Consequently, the selection and rotation phases do not have to be disjointed. Three scenarios arise:

- The selection phase and the Graz protocol are sequential and disjoint (later in the text: Graz v1)
- The selection phase starts during the instructions of the Graz protocol (later in the text: Graz v2)
- The selection phase starts when the left/right feedback from the headset starts being displayed (later in the text: Graz v3)

In this preliminary experiment however, we only consider the evaluation of Graz v1, while Graz v2 and

Experiments

In order to evaluate the modality combination, we performed experiments on 30 participants aged between 22 and 42 years old, among whom were 5 women and 25 men.

All the participants were complete novices and had never used BCIs beforehand, and were all in good health.

For each modality, fatigue is a factor we fear would constitute a strong bias that would be very apparent if we chose a fixed order for the evaluation of the three combinations. Thus, we split the population in three groups that each used the combinations in different orders: (A) Graz + SSVEP + Eye Tracker, (B) SSVEP + Graz + Eye Tracker, (C) Eye Tracker + Graz + SSVEP.

The training phases are systematically performed just before the modality is used.

Furthermore, we had four trials per modality combination.

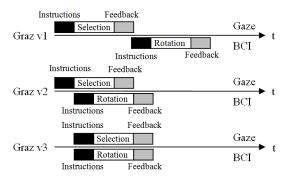


Figure 4. Possible combination of Eye Tracking and Graz Motor Imagery

Qualitative Criteria

After the end of the experiment for each of the subject, we informally ask the opinion of the subjects about how they felt regarding the various modality combinations, what aspects hindered or favoured the interactions, what they liked and did not like.

At the end of all the experiments, we defined categories for common answers and counted how many users were of a similar opinion.

Quantitative Criteria

We are primarily interested in how many errors the users make during the selection and rotation phases.

This is why we keep track of the number of correct/incorrect selections as well as the number of correct, incorrect and total rotations. These values allow us to calculate the error rates for the selection and rotation tasks.

Furthermore, we are interested in evaluating the efficiency of the combinations of modalities: how long it takes on average to select or rotate a tile. Therefore, at each turn, we keep track of the time it takes to select the correct tile and then the time it takes to make the

correct rotation. Consequently, we expect errors would result in higher selection and rotation times.

Fatigue being an important factor in the use of BCI and of eye tracking (to a lesser extent), we planned four trials for each modality, which enables us to track the changes of the error rates and timings and thus, are able to evaluate the effect of fatigue on performance.

Interaction Modalities

Before analyzing the results of the experiments we feel that it is useful to present each of the modalities in more depth so as to give the reader a better understanding of the interpretation of the results.

Eye Tracking

The calibration is done using the calibration programme provided with the Tobii SDK.

We use the centroid of the gaze coordinates to determine the point where the user is looking. We used a short dwell time of 200ms so as to enable a fast interaction while still mitigating the errors due to noise.

Graz Motor Imagery

Given the lack of electrodes of Emotiv Epoc over the motor cortex and the low resulting classification accuracies, for the experiments involving Graz, we placed the Emotiv headset tilted towards the back of the head. We shifted the electrodes FC5 and FC6 to C5 and C6; the F3 and F4 to FC3 and FC4; AF3 and AF4 on F3 and F4 (**Figure 1.b**).

The idea behind the Graz protocol is to imagine right and left hand movement to control an interface through a feedback value corresponding to the degree of left or right classification.

For the acquisition we need to capture signals of when the users imagine left had movement and of when they imagine right hand movement. The captured signals then needs to be transformed into feature vectors for the classifier training. The training consists of 20 trials for each class, intertwined in a random order. For each training trial, a cross first appears on the screen to focus the user's attention, followed 1s later by an arrow (displayed for 1.25s) that instructs the users what movement to think about. From the moment the arrow is displayed until the cross disappears (2.75s).

For the training, the signals are first filtered by a reference channel (to filter background EEG), after which they go through a spatial filter (Surface Laplacian) that aggregates the signals into only two channels. Then, the signals are separated into epochs that correspond to left and right hand stimulations and further into 1s epochs with a 0.062s interval. Subsequently, the epochs are squared, go through a moving average filter and are finally scaled using a log(1+x) filter. Each epoch becomes a single feature of the feature vector, so as to obtain separate feature vectors for each class that are fed to the classifier.

On average for Graz we had **61.3%** for the classification accuracy with a standard deviation of **4.3%** using 5-fold cross validation.

For the online use, we directly acquire the signals from the headset. The users are shown a cue telling them to start thinking about the desired hand movement. The signals go through the same signal-processing pipeline as for training and lead to feature vectors (no labels) that are fed to the classifier.

SSVEP

For the needs of our test application, we require two different targets and thus two flickering frequencies. The screen we used for the experiments has a refresh rate of 60 Hz with which the maximum flickering

frequency possible is 30 Hz. Furthermore the frequencies have to be divisors of the maximum frequency (30Hz). We chose 15 Hz and 30 Hz.

We need a CSP filter (Surface Laplacian) that aggregates the signals from all the channels into only two, by applying weights on each channel that correspond to their respective saliences.

For each flickering target, the signals of that target are separated from the rest of the classes. The former are considered as one class (positive examples) for the CSP training and the latter are considered as a second class (negative examples). All of the signals are then filtered through a Butterworth band pass filter around the corresponding frequency (for example between 14.625 and 15.375 for the left target). The resulting signals are then aggregated into epochs corresponding to the stimulation duration (7s) and then used to train the filter.

For the training of the classifier (LDA) the special filter previously trained is applied, followed by the same target separation and filtering procedure as described above for the training of the filter.

Subsequently, the signals undergo the same transformation into feature vectors as for the Imagined Hand movement implementation

For each target, we obtain feature vectors for positive examples and negative examples that can be used to train a classifier for that particular target.

For the online use, the signals are filtered with the spatial filter and then fed in parallel to both the trained classifier. First they are filtered around the appropriate frequency (+/- 0.375 Hz) with a band pass filter. Subsequently, they are transformed into feature vectors following the same process as for Graz, and sent to the classifiers.

Using five-fold cross validation, on average the classification accuracy for SSVEP was **79.8%** with a standard deviation of **1.3%**.

Results

We will first describe the qualitative results stemming from the informal interviews of the users followed by a quantitative analysis according to the metrics described earlier.

Qualitative analysis

The users unanimously agreed that selection through gaze was very fast and intuitive; however 8 people (26%) felt that the controls for the rotation operations felt unnatural and uncomfortable.

For Graz, while some users pointed out it were a very interesting modality, they all complained that the process was slow and tedious and that the classification often seemed random. Some people (60%) found it to be a very natural interaction while 73% tedious and tiresome.

Users generally found SSVEP to work better than Graz and to be much easier to use (76%). However, they found the training phase more tedious, and in general people are disturbed or tire quickly due to the flickering. They unanimously expressed the wish for the rotation phase to be shorter.

In general the users felt that the pure eye tracking approach was the easiest and least tiring, while admitting that BCI interactions are interesting. Some

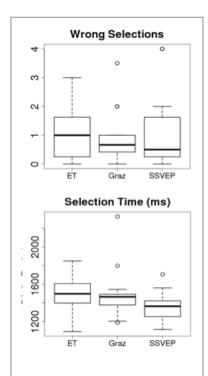


Figure 4. Box plots of the results for the number of wrong selections and the average selection time.

people suggested that BCIs could be better suited as a complementary and optional modality.

Modality	Wrong	Wrong	Avg. Sel.	Avg. Rot.
Combination	Selections	Rotations	Time. (ms)	Time (ms)
Pure E.T.	0.930	2.00	1486	668
	σ0.524	σ0.612	σ170.0	σ98.9
E.T. +	1.328	3.992	1477	3040
Graz	σ0.619	σ0.307	σ147.9	σ301
E.T. +	1.000	1.717	1367	1161
SSVEP	σ0.474	σ0.825	σ117.4	σ261.9
p-value	0.8426	<0.001	0.0845	<0.001

Table 1. Number of Wrong Selections, Rotations, Selection time and rotation time (average and standard deviation)

Quantitative analysis

Before analysing the results we wanted to ensure the statistical significance between the observed groups for each of the metrics we considered, which led us to use a one-way ANOVA for each of the metrics and between all the groups, which is based on the assumptions of independence, normal distribution of the data and homogeneity of variances.

The first assumption is met, considering that the evaluation of one modality combination did not directly influence the other modalities (assuming that the effects of fatigue are homogeneous between the groups). In order to verify the normality of the distribution, we used normal Q-Q plots and computed the correlation between the theoretical (normal) and

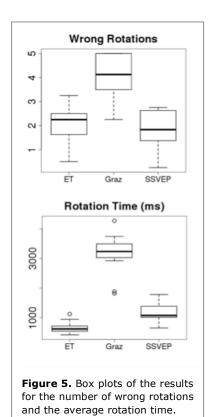
empirical quartiles (equivalent to a Shapiro–Wilk test). The correlation values obtained for each of the groups were all above 0.95 which indicates a significant correlation with the normal distribution, by rejecting the null hypothesis that the distributions are not normal (p>>0.05). We checked the assumption of the homogeneity of variances using Levene's Test, and obtained p values below 0.05, thus indicating a significant homoscedasticity.

Furthermore, whenever we witnessed a significant p value between the groups for a given metric, we applied a Tukey HSD post-hoc pair wise test, in order to determine the significance between all possible pairs of groups at a 0.05 level.

In **Table 1** are summarised: the number of correct and incorrect selection and rotations, the average rotation and selection times for each combination of modalities. **Figure 4** and **Figure 5** show box plots of the distributions of results for each of the metrics and modality combination. Since we systematically had significant post-hoc p-values between the groups when the p-values of ANOVA were significant, we only indicated the ANOVA p-values, which correspond to the maximum post-hoc p-value among all the group combinations.

For both the selection errors and times, any variation in the values is not significant (p >> 0.05).

In terms of wrong rotations, the variations are significant (p<0.05) and SSVEP is the best, followed by the Pure eye tracker, which has a higher variation; while Graz is notably worse with about 2 wrong rotations more than the two others. This seems to confirm that the users found the rotation method with the pure eye tracker unnatural and uncomfortable, whereas the SSVEP approach was more straightforward.



Lastly, the rotation times, which also exhibit significant variations (p<0.05), reflect the comments of the users, that rotating with Graz is difficult, leading to notably higher rotation times and rotation time standard deviation with up to 3 seconds more than the other modalities.

The lowest rotation time was with the eye tracker, which despite more errors, allows for faster rotations. SSVEP is somewhat slower with on average 1.3s more per rotation.

Conclusions

Using BCI in a gaming context with non-disabled people is a challenging task, as it puts a lot of constraints on usability and available technologies. In our work, we attempted to compare several combinations of BCI and eye tracking. After performing our experiments for the evaluation of three modality combinations: Eve tracking only, Eye tracking and Motor Imagery (Graz), Eye tracking and SSVEP. We found that the eve tracking only was generally the less error prone and quickest modality combination. Of course, for the experiments we used a high quality eye tracker and it is not very surprising that the pure eye tracking solution would reach better results. However, eye tracking was certainly not the most natural way of performing a rotation task for the users. When using BCIs, SSVEP manages to lead to a performance very close to pure eye tracking, which is commendable but still is perceived as a not too natural way of performing a rotation. As for the Motor Imagery Graz protocol, it was the most natural interaction modality for the task at hand according to the users. However, the performance greatly suffered from the low classification accuracy that led to many errors and much slower interactions. It is apparent that even with the shift of the Emotiv Epoc headset, the coverage of the motor cortex was still insufficient to provide a usable alternative to SSVEP and a good quality eye tracker.

The Motor Imagery modality would greatly benefit from using better EEG acquisition equipment. One possibility would be to exploit medical-grade amplifiers such as the g.tec USBAmp or alternatively to consider other consumer grade devices with a better motor cortex coverage. As a first step towards exploring these possibilities we are already in the process in performing more extensive experiments with a g.tec USBAmp. On top of using a better EEG acquisition device, we will also include P300 is the considered modalities for selection in order to have a pure BCI system as well as a baseline.

References

- [1] Cipresso P., Meriggi P., Carelli L., Solca F., Poletti B., Lulé D., Ludolph A.C., SilaniV., Riva G. 2012. Brain Computer Interface and Eye Tracking for Neuropsychological Assessment of Executive Functions: A Pilot Study In: Proceedings of the 2nd International Workshop on Computing Paradigms for Mental Health pp.: 79-88 SciTe Press
- [2] Coutaz J., Nigay L. Salber D., Blandford A., May J., Young R., Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE properties, ProcINTERACT'95, Chapman & Hall Publ., Lillehammer, Norway,1995, 115-120
- [3] Drewes H. 2010. Eye Gaze Tracking for Human Computer Interaction. PhD Thesis. LFE Medien-Informatik der Ludwig-Maximilians-Universität
- [4] Emotiv Epoc Website Available: http://www.emotiv.com/
- [5] Gürkök H., Hakvoort G., Poel M. 2011. Modality switching and performance in a thought and speech controlled computer game. In Proceedings of the

- 13th international conference on multimodal interfaces (ICMI '11). ACM, New York, NY, USA, 41-48
- [6] Hachmeister N., Riechmann H., Ritter H., Finke A. 2011. An approach towards human-robot-human interaction using a hybrid brain-computer interface. In Proceedings of the 13th international conference on multimodal interfaces (ICMI '11). ACM, New York, NY, USA, 49-52.
- [7] Jacob R.J.K. 1990. What you look at is what you get: eye movement-based interaction techniques. In Proc CHI '90, Jane Carrasco Chew and John Whiteside (Eds.). ACM, New York, NY, 11-18.
- [8] Lalanne D., Nigay L., Palanque P., Robinson P., Vanderdonckt J., and Ladry J.F. 2009. Fusion engines for multimodal input: a survey. In Proc. of the 2009 international conference on Multimodal interfaces (ICMI-MLMI '09). ACM, New York, NY, USA, 153-16
- [9] Middendorf M., McMillan G., Calhoun G., Jones K.S. 2000. Brain-Computer Interfaces Based on the Steady-State Visual-Evoked Response. IEEE Transactions on rehabilitation Engineering, Vol. 8, No. 2, June 2000
- [10] Neuper C., Müller-Putz G.R., Scherer R., Pfurtscheller G. 2006. Motor imagery and EEGbased control of spelling devices and neuroprostheses, In: Christa Neuper and Wolfgang Klimesch, Editor(s), Progress in Brain Research, Elsevier, Volume 159, Pages 393-409
- [11]Nijholt A., Allison B. Z., Jacob R.J.K. 2011. Brain-computer interaction: can multimodality help? In Proceedings of the 13th international conference on multimodal interfaces (ICMI '11). ACM, New York, NY, USA, 35-40.
- [12] Renard Y., Lotte F., Gibert G., Congedo M., Maby E., Delannoy V., Bertrand O., Lécuyer A. 2010. Openvibe: An open-source software platform to

- design, test, and use brain--computer interfaces. In real and virtual environments. Presence: Teleoper. Virtual Environ. 19, 1 35-53.
- [13]Sundstedt V. 2010. Gazing at games: using eye tracking to control virtual characters. In ACM SIGGRAPH 2010 Courses (SIGGRAPH '10). ACM, New York, NY, USA, Article 5, 160 pages.
- [14]Vilimek R., Zander T. O. 2009. BC(eye): Combining Eye-Gaze Input with Brain-Computer Interaction. In Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environnements (UAHCI '09), Constantine Stephanidis (Ed.). Springer-Verlag, Berlin, Heidelberg, 593-602.