Conceptual Priming for In-game BCI Training

NATALIYA KOSMYNA, Grenoble INP
FRANCK TARPIN-BERNARD, Univ. Grenoble Alpes
BERTRAND RIVET, Grenoble INP

Using Brain Computer Interfaces (BCIs) as a control modality for games is popular. However BCIs require prior training before playing, which is hurtful to immersion and player experience in the game. For this reason, we propose an explicit integration of the training protocol in game by a modification of the game environment to enforce the synchronicity with the BCI system and to provide appropriate instructions to user. We then dissimulate the synchronicity in the game mechanics by using priming to mask the training instruction (implicit stimuli). We conduct an evaluation of the effects on game experience compared to standard BCI training on 36 subjects. We use the game experience questionnaire (GEQ) coupled with reliability analysis (Cronbach's alpha). The integration does not change the feeling of competence (3/4). However, flow and immersion increase sizably with explicit training integration (2.78 and 2.67/4 from 1.79/4 and 1.52/4) and even more with the implicit training integration (3.27/4 and 3.12/4).

Categories and Subject Descriptors: I.5.5 [**Implementation**]: Interface System; H.5.2 [**User Interfaces**]: Input Devices and Strategies

General Terms: Brain Computer Interfaces, Training In-game, Co-learning

Additional Key Words and Phrases: Interactive machine learning

ACM Reference Format:
Nataliya Kosmyna, Franck Tarpin-Bernard, and Bertrand Rivet. 2015. Conceptual priming for in-game BCI training. ACM Trans. Comput.-Hum. Interact. 22, 5, Article 26 (October 2015), 25 pages.
DOI: http://dx.doi.org/10.1145/2808228

26

1. INTRODUCTION

BCIs allow capturing and recognizing the brain activity of users with the purpose of generating commands for any computer system. There are many applications of BCIs, traditionally in medicine (interaction with locked-in patients, typing) [Wolpaw et al. 2002], but also increasingly in out-of-the lab applications such as: control (e.g. car, robotic arm) Chapter 6.2 in [Tan and Nijholt 2010]; user-state monitoring (e.g. workload) for interface and task adaptation; as well as gaming and entertainment.

There are three types of BCIs: *passive*, in which the BCI monitors and detects the state of the user (e.g. workload monitoring); *reactive* in which external stimuli are presented to the user and in which the resulting activations are recognized (mostly for speller applications); *active*, in which users voluntarily perform an imagined mental

Authors' addresses: N. Kosmyna, Bureau B205, Bâtiment IM2AG B, Laboratoire d'Informatique de Grenoble, 41 Rue des Mathématiques, BP 53 38041, Grenoble CEDEX 9, France; email: natalie@kosmina.eu; F. Tarpin-Bernard, Bureau B203, Bâtiment IM2AG B, Laboratoire d'Informatique de Grenoble, 41, rue des Mathématiques, BP 53, 38041 Grenoble cedex 9; B. Rivet, Bureau D1122, 11 Rue des Mathématiques, Domaine Universitaire, BP 46 38402, Saint Martin d'Hères Cedex, France; email: bertrand.rivet@gipsa-lab.grenoble-inp.fr.

DOI: http://dx.doi.org/10.1145/2808228

action and in which the resulting activity is recognized (e.g. direct control applications). In this work we focus on active BCIs. A *crucial* component for active BCIs is ability of users to produce stable brain signal patterns in order to facilitate their recognition. Here, we define this skill as "BCI *training/learning*". *BCI training* could require hours of training and repetitive practice, especially for the continuous use of a BCI system. Nowadays, acquiring this skill still remains a challenging task because of the following reasons:

1. *BCI training is not an interactive process.* It involves users by asking them to successively repeat the imagination of an action (for example, imagining moving the hands for Motor Imagery (MI) based BCIs) following stimuli synchronized with a fixed timing (a.k.a. synchronous BCIs). For example, the user can only perform the action every 10s when the system displays a cue marking the start of a trial period. Such a setting tends to be unappealing and to increase the level boredom in users. One of the possible solutions consists in using BCI systems that allow users to interact asynchronously: the classification occurs at short discrete intervals (e.g. every 250ms), thus allowing users to interact freely. Still, the synchronous setting remains common, as the performance of such BCIs can easily be evaluated, thus making this setting desirable for experiments and for comparing system in an in-the-lab setting. Moreover, a continuous classification (required for asynchronous systems) greatly increases the computational requirements towards achieving a real-time BCI system.
2. *BCI training is a long process.* On the one hand, BCIs based on *the machine learning approach* require, on an average, 10 to 20 minutes of training for 2 actions to be recognized with an accuracy of 80% [Wolpaw et al. 2002]. On the other hand, another technique known as *operant conditioning requires* training sessions that can last up to several months (for around 90+% accuracies) [LaFleur et al. 2013]. However, compared to the inexistent training required for pervasive tablet or keyboard based interactions, training a BCI system is a tedious process.
3. *BCI training is not an engaging process.* As mentioned previously, most BCI systems are based on a machine learning approach, in which *BCI learning* by users remains incidental. Consequently, only the classifier is being explicitly trained and *BCI learning* in users is not emphasized. Users don't have an easy way of knowing whether they imagine the action in a consistent or correct manner, so as to enable the system to recognize their brain signal patterns better. One solution to alleviate the severity of this issue is to provide *feedback* about the classification process and particularly, neurofeedback that consists in displaying the user's brain signals back at them (Figure 1). This allows users to get a sense of how to modulate their brain signals during the training of the classifier, which contributes to making the job of said classifier easier. Yet, even nowadays, feedback is designed at a low level and is not easy to understand and to interpret by users. Moreover, it relies on unappealing visualization representations. Currently, the improvement of feedback in BCIs is an essential step towards improving BCI performance and their usability.

One approach towards solving these issues is to propose improved signal processing techniques (e.g. through Riemannian geometry [Congedo et al. 2013]) and classifiers in order to make training faster, improve performance and to require less training data overall. Another approach is to take interest in human learning and to improve training protocols and feedback strategies in order to provide a more enticing experience to users and to enable a wider range of users to use BCIs. While both aspects are equally important, the former has garnered most of the research efforts [Lotte et al. 2013].

As for the latter, the improvement of training and feedback protocols, efforts have been focused on either providing biased feedback in line with instructional design practices (positively biased feedback improves performance in novice users [Barbero and

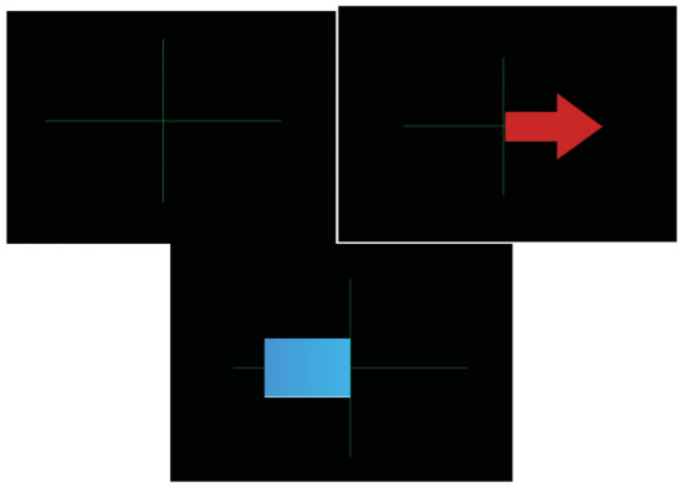


Fig. 1. Example of feedback in traditional BCI training protocols (usually for expert users). Classification feedback: an arrow pointing right indicates to the learner to imagine a right hand movement. The blue bar gives feedback with its direction and length based on the classifier output. Thus, it shows how well the mental task is recognized. The bar extends towards the left for an imagined left hand movement, and toward the right for the imagined right hand movement. (http://OpenViBE.inria.fr).

Grosse-Wentrup 2010; Vidaurre et al. 2010]) or in building more enticing interactive environment, for example using virtual reality (VR) or video games [Lécuyer et al. 2008]. There have been many works using BCIs for the purpose of controlling games (e.g. controlling World of Warcraft [Bos et al. 2010], "Brain Invaders" [Congedo et al. 2011], serious games [Sung et al. 2012], research games [Kosmyna and Tarpin-Bernard 2013]).

While controlling games or VR during the use of the BCI creates a more immersive experience, the training is typically just as tedious: it remains separate from real-time use within the environment. The next step in further coupling BCIs with games and with VR seamlessly is to integrate the training phase in the environment. Contrarily to games, VR environments are general purpose and do not offer a convenient way of integrating the training phase without the presence of game mechanics and storyline elements. Flatla et al. [2011] propose a general framework and guidelines for the design of games around calibration protocols for various interaction modalities that require calibration prior to use. For each type of calibration, they explain what game mechanic to adapt and how.

The first limitation of the work is that BCI training does not entirely overlap with traditional calibration types. BCIs that are based on machine learning classifiers need to capture signals for several categories of brain activity synchronously, which requires more complex gameplay and that specific integration mechanics be put in place. Moreover, the games in Flatla et al. (2011) are constructed for the purpose of performing the calibration. The calibrated input modalities are then used on tasks that have no relation to the game. In other words, the calibrated modality is not meant to control the game. For machine-learning based BCIs that are meant to be used as control modalities for an existing game (i.e. not crafted specifically for the BCI), a training integrated in said game can be beneficial in order to improve the training. Indeed, BCI learning is a form of human learning and inspiration can be drawn from the design of educational games and the framework of formative feedback [Lotte et al. 2013]. Lotte et al. [2013] particularly cite immersive virtual environments as a way or improving BCI training.

The main challenges lie in the fact the BCI training is synchronous and that a way must be found to integrate in the environment and then to further mask the training in the game mechanics to make it seamless to the player.

If the objective of integrating BCIs in games is immersion, then the nature of the BCI control must correspond to the semantics of the task. If the objective is to move a character for example, imagining movement is a good match in terms of BCI control. On the other hand if the objective of the BCI were to trigger discrete actions, then it would be ideal for the user to just visualize or conceptualize something related to the action to trigger. Let's say that we want to use the BCI to switch between weapons in a shooter game, then imagining a weapon would be the type of control that best match the semantics of the task. Some type of BCIs can trigger discrete actions through visual stimuli (e.g. flickering target), however such a control is independent from the semantics of the task and is less immersive. This is why, in this paper, we focus on BCI systems based on conceptual and mental imagery [Simanova et al. 2010].

One way of integrating synchronous BCI training is to directly transpose it in the game environment by ensuring that players follow a progression that is consistent with the state of the BCI system (we call this synchronous in-game training or IGTs+). The game should give instructions for each class and allow players to trigger the beginning of the training for a particular class. When one class is trained, the game should let the player move forward to another section for the training of another class. Although this leads to an integrated training, the training still does integrate with the gameplay and does not match the semantics of the game, it is merely a stage required before playing.

The next step in the integration of training is to mask the synchronicity of the training within the game mechanics so that players do not realize that they are training a BCI (we call this in-game training, implicit or IGTi). In order to do this, we need to make sure the user imagines the right concept at the right time. In other words, the player must be "primed" by some stimulus that leads them to imagine the desired concept. Priming is an implicit memory effect in which the exposure to a stimulus influences the user's response to another [Schvaneveld and Meyer 1973]. Conceptual priming in particular is a type of priming in which the stimulus is the image or shape of an object with the effect of reinforcing the effects of stimuli pertaining to related objects in shape or appearance[Biederman and Cooper 1992].

More concretely, in the present paper we integrate the BCI training seamlessly into a game and we want to detect the mental imagery of concepts such as "weapon" or "flashlight" (that are relevant to the particular game to which the system is applied) as we'd like to trigger corresponding in-game actions.

We implement both synchronous in-game training (IGTs+) and implicit in-game training (IGTi+) in the First Person Shooter (FPS) game Doom 3 in order to evaluate and compare their effect on game experience (immersion, flow, and competence) with relation to a standard BCI baseline trained outside of the game.

Let us now summarize the outline of this article. We first introduce background and related work. Then we provide the general guidelines we used to design a BCI system within the game and how we implemented them in practice. Subsequently we described how we implemented our BCI system, the modification of the original game to make BCI input and communication possible and the integration of the BCI in the game. We move on to the experiments, the protocol we followed as well as the results and their discussion. Finally, we conclude and discuss about possible future applications of our work.

2. BRAIN COMPUTER INTERFACES (BCI)

A BCI system is a closed loop between a user and the system. Generally, the user interacts with the system and the system gives feedback about its decision state after

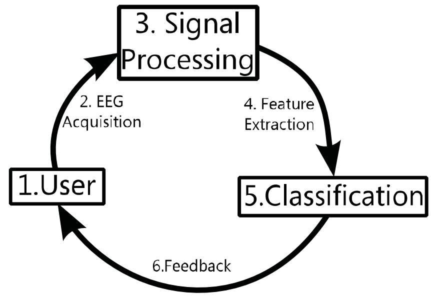


Fig. 2. The common BCI loop.

the interaction. The BCI loop is composed of (in order): the signal acquisition process; then the signal processing step in which signals are processed and prepared; feature extraction phase in which we identify and extract salient features from the signal; the classification stage in which the features are matched to the classifier model to identify one of the phenomena it was trained to capture (Figure 2).

More formally, the objective of a BCI system B is to assign a brain signal S_t of a fixed duration (an *epoch;* e.g. 1s) at time t, to a class Cl_i from a set of N classes Cl that correspond to a set of brain activity states BS_i that need to be recognized.

A machine learning classifier C is trained to recognize the desired states BS_i through a set of *training examples* $T(Cl_i)$ for each class Cl_i. A training example is a signal epoch of the same duration as S_t that was recorded when the user was in the desired state BS_i corresponding to the class Cl_i. This is called the training phase.

C returns, for each S_t, a classification outcome C_o. C_o is a map that associates to each C_i, a confidence weight or posterior probability w_i. Note that not all classifiers give a distribution across outcomes, for example the output of Linear Discriminant Analysis (LDA) is the label of the selected class. However, we can see the latter as a particular case in which the probability associated to the selected class is 1 and the probability associated to the other classes is 0. The element of C_o with the highest weight is the most likely assignment for S_t. While the system gives classifier feedback to the user, another essential element is to give neurofeedback to the user that shows a representation of desirable features extracted from user signals and that allow users to learn how to modulate their brain signals.

2.1. BCI Paradigms

When the set of brain states to recognize BS_i are the result of a conscious action of the user, the BCI is called *active*. When the BS_i are the result of an external stimuli, the BCI is called *reactive*. When the BS_i are the result of a passive state of a user, the BCI is called *passive*. In this article, we only deal with active BCI systems. The most common active and reactive BCIs include: MI and Conceptual Imagery [Simanova et al. 2010].

Most paradigms are based on event-related potentials (ERP), an activation in a certain area of the brain in response to a stimulus (event). In general, ERP are either positive or negative. The name given to ERPs often starts by P or N depending on whether they are positive or negative, followed by a number in milliseconds that characterizes how much time the potential appears after the stimulus. The main BCI paradigms are the following:

- Motor Imagery (MI) – the user imagines moving various body parts, for example hands, feet, tongue [Blankertz et al. 2008].

- Steady State Visually Evoked Potentials (SSVEP): the user looks at a target flickering at a certain frequency, for example 15Hz. This causes a rapid succession of action potentials in the visual cortex, some of which are at the same frequency as the stimulation [Kelly et al. 2005].
- P300: P300 is a positive action potential generated 300ms after the user makes a choice (conscious or otherwise). By putting a set of items to choose from in a grid or matrix and by successively flashing the items, it is possible to trigger a P300 when the item the user wants to select flashes [Wolpaw et al. 2002].
- Error potential based BCIs (ErrP) exploit error related negativity (ERN), ERPs that are negative activations generated in the brain 150ms after the stimulation onset when the user commits and error (even if not consciously aware) or when negative feedback is received [Iturrate et al. 2010]. ERNs are often used to produce adaptive BCI systems that can detect when the user perceives a classification error and adapt the classifier in accordance [Thomas et al. 2013]. They also have important applications in interaction design, in which they can serve to detect desirable properties of interactive processes, without explicit feedback from users [Vi et al. 2014].
- N400 are another category of error-related potentials that manifest as a negative activation 400ms after the stimulus onset. N400 is part of the normal brain response to words and other meaningful (or potentially meaningful) stimuli, including visual and auditory words, sign language signs, pictures, faces, environmental sounds, and smell [Kutas and Federneier 2000].
- Conceptual imagery is a recent development, in which the aim is to capture EEG in order to detect when the user thinks of a category of conceptual objects (hammer/tool, cat/animal) by Simanova et al. [2010]. They considered textual, visual and auditory cues. They find that the best performing BCI system resulted from visual stimuli. The stimuli featured abstract or concrete images of objects from various categories that the BCI proceeded to recognize. Thus, conceptual imagery constitutes a powerful type of BCI that can lead to more natural interactions for users by using the semantics of the task. For example, one could imagine/visualize a key to trigger the ignition of a car. Regardless of the type of imagery task, a crucial step in the creation of a BCI system is the training of the classifier that allows to recognize the various desired brain activity states but also to train the users to modulate their signals. The first step in this process is the capture of the signals from users.

2.2. Towards Task Specific BCI Training

BCI training can also be categorized in synchronous versus asynchronous. A synchronous training protocol displays cues to the user at fixed timings to indicate to the user what imagined action to perform or what stimulation target to focus on. In asynchronous training on the other hand, no cues are displayed to the user. However, in order to train a classifier, it is required to have a synchronous training protocol given that we need to capture signals belonging to each class of brain activity to recognize. Without synchronous cues we would essentially have to rely on a completely unsupervised BCI systems, in which it would be very difficult to separate signals for each activity. For this reason, almost all BCI systems make use of a synchronous training session to train the classifier and then can potentially use an asynchronous training phase to tune the parameters and thresholds of the system for a specific subject (e.g. [Rohm et al. 2010], in which an asynchronous training phase is used after the standard synchronous phase to tune subject specific parameter).

The requirement for synchronous training means that it will be difficult to make the training specific to a particular application in a seamless way, as synchronous interaction is not very natural. As a consequence, BCI training sessions (and feedback) are typically task independent. The positive consequence is that the BCI can be used in several different applications and tasks without the need for re-training. However, the negative consequence is that a task specific training may have yielded a better performance and if the BCI is intended to serve a single specific task the gain in generality of task-agnostic training is lost.

Another issue with the requirement for synchronicity is one of the main reasons for the tediousness of BCI training phases. Attempts have been made to train BCI systems incrementally instead of having a one-off training phase and to reduce the amount of training required through improvements in filtering and signal processing techniques. Incremental BCI training implies that the training and the online use are interlaced. An initial shorter training phase or a signal database is used to initialize the classifier and new training examples are added as the BCI is used. Acquiring the subsequent training example still requires synchronicity; however, diluting their capture makes the training more natural and pleasant. Moreover, incremental BCIs have been found to exhibit better performance [Millán and Mouriño 2003]. As such, incremental BCI training is more suitable for asynchronous BCI systems. The reason for the increased performance is that during online use, through task-related feedback, users learn to better modulate their signals and the modulated signals are taken into account in the training of the classifier.

In this work, we take a particular interest in the integration of BCIs and their training with video games. Video games and VR provide a more immersive and motivating environment for users and are synergetic with BCIs in several ways [Lécuyer et al. 2008]. The more immersive experience improves the focus and concentration of users and lead to a better BCI training and the BCI allows adapting the game to the user and making the experience even more seamless. To our knowledge, there have been no attempts to integrate BCI training proper in games. Asynchronous BCI training corresponds to the notion of calibration that is required for many interaction modalities (gamepads, eye trackers, prosthetics). For BCIs that only monitor brain activity passively in which the objective is to capture signals to determine the user state and for simple reactive BCIs that detect the brain's response to external stimuli (for example, Visual Evoked Potentials or VEP) the only training required is to determine the activation threshold for each particular user and thus no synchronicity is required.

The main challenge remains the seamless integration of synchronous BCI training with games both for training and for control in a fully immersive and incremental manner.

This challenge involves and requires the development of two aspects:

- Incremental BCI systems require a constant interaction with users, as they need to iteratively capture signals to improve the training of the classifier and to perform adaptations of the BCI system. Given that the system also gives feedback to the user about the classification and the features captured, there is a bidirectional feedback loop between the user and the system (notion of co-learning BCIs [Kos'myna et al. 2013]) and is an application of the more general domain of Interactive Machine Learning (IML) and its integration with games (Section 3).
- The synchronicity constraint and explicit training cues are cumbersome and hinder immersion and the seamless integration of BCI training. Strategies must be explored that can make the cues implicit and integrated seamlessly in the game, for example, by the use of priming (Section 4).

3. COMBINING GAMES AND BCIS: INTERACTIVE MACHINE LEARNING

Co-learning requires a feedback loop between the user and the system and corresponds to the notion of IML: the system gives feedback to the user (classifier visualization of outcomes or features) and the user then gives feedback to the system (new parameters, selection of features or training instances for the classifier) [Fails and Olsen Jr. 2003].

As far as co-learning BCIs are concerned, we are interested in the particular case of real-time IML with supervised classifiers.

If we examine first work to introduce the notion of IML, [Fails and Olsen Jr. 2003] and the work of Talbot et al. [2009], one of the first applications of IML (Ensemble-Matrix) to a HCI, three main use cases for IML for supervised machine learning can be derived, [Fiebrink et al. 2011]:

- Interactively editing training data;
- Supporting the use by nonexperts;
- Correspondence to real user priorities.

The interaction of the user with the classifier is typically done through GUI controls, other examples include [Kapoor et al. 2010], in which an interface is proposed to set classification weights by directly modifying the confusion matrix. The other important element of IML is the design of an effective classifier visualization method; however contrarily to our in-game application IML visualization is generally task – agnostic and explicit (as opposed to specific to the game and implicit through the use of priming). Moreover, we are mainly interested in the requirement for IML that enable the use by nonexperts and that mandate the correspondence to real user priorities, as the main objective is the play, the game as the interactive edition of training data is incompatible with the implicit nature of a priming-based training.

4. PRIMING

Priming is known to be subliminal, if it occurs within a very short period of the time (e.g. around 20 to 50ms) and it is usually followed or preceded by a mask (random geometrical figures or dashes that conceal the priming stimulus) to reduce or eliminate conscious perception. This kind of priming is also known as masked priming.

Supraliminal (visible) priming [Greenwald et al. 1996], on the other hand, could be projected for longer periods of time (up to 400ms or even longer) and is not masked with any additional projections. Denning et al. [2011] performed an experiment with supraliminal priming and found that the effects of the priming were still present 14 and 21 days respectively after the subjects were primed.

Priming types can also be categorized depending on what processes in the subjects are targeted by the priming: cognitive (such as reasoning or decision making) or affective (eliciting some emotional reactions for for example [Lewis et al. 2011] on a creativity task or [Harrison et al. 2013] on influencing graphical perceptions).

Ruijten et al. [2011] assumed that when people are primed with a concept that unconsciously activates a certain goal, they are better able to use information embedded their environment to reach the primed goal.

Words, sounds, images or objects could be used as primes. We will now review some relevant work in areas in which priming was already applied that are close to our research interests, as mentioned in the introduction, notably, priming in games, priming and learning and finally priming for BCIs.

4.1. Priming in Games

Although the priming was applied in different domains, it got a lot of success particularly in games. Resent research [Green and Bavelier 2003, 2006a, 2006b, 2007; Li et al. 2009; Strohbach et al. 2012] reveal that action video game players outperform nonplayers in a wide range of attention, perceptual, and cognitive tasks.

Graybill et al. [1985] showed that the children who played violent games were more assertive than nonviolent video game players. Schutte et al. [1988] found that children who played violent video games were more likely to be aggressive than others. All these examples could be generalized with a famous «weapons priming effect», described by Berkowitz and Lepage [1967]. The researchers show that stimuli commonly associated with aggression (like weapons) can elicit more aggressive responses from people «ready to act» aggressively.

Pohl et al. [2014] suggest as well that the action video games support shortly presented visual stimuli. Regarding all this information we consider action game a nice platform to test the priming.

4.2. Priming and Learning

The work of Chalfoun and Frasson [2012] aims at producing a novel Intelligent Tutoring System (ITS) that introduces cognitive primes to implicitly enhance reasoning in students in problem solving environments (as opposed to explicit and direct strategies employed in other ITSs. They argue on the basis of experimental evidence in neuroscientific studies that learning is interplay of conscious and unconscious processes and that both need to be studied to fully understand and maximize learning in students. Their idea is to project the answers of a related problem under the visual threshold of learners (in a way that does not disturb the learning task) while they are actively undergoing the learning process. They apply their system on a learning task in a 3D environment, in which learners and teaching entities are represented like agents. They find that their technique enhances unconscious learning without affecting active conscious learning negatively. Based on neural recording during their experiments, they hypothesize that cognitive priming allows not only assessing conscious reasoning but also intuitive reasoning.

4.3. Priming and BCIs

Priming has been mainly applied to BCIs for the elicitation of N400 activation, first introduced in the work of Kutas and Hillard [1980]. In most of the literature, subjects are primed by an explicit stimulus (word or picture) while their signals are recorded to identify the N400 activation. They are then shown a second related stimulus (*probe* stimulus) that allows verifying whether the probe stimulus leads the generation of a similar N400 activation. They find that the amplitude of the N400 activation is modulated by the semantic/associative relatedness of the prime and probe stimuli. Among numerous research investigations that aim at studying N400 in more detail, two of them are of particular interest. First, the work of van Vliet et al. [2010] aims at achieving an N400 activation without showing any priming stimuli. His results indicate that subjects can prime themselves on a physical object by actively thinking about said object during the experiment. This opens the door to the possibility of BCIs that can be controlled by simply thinking about a word. Moreover, van Gerven et al. [2013] have concluded that it is possible to detect N400 at the single-trial level, although the classification accuracies are low.

5. OUR BCI SYSTEM

The architecture is based on Minimum Distance Classification (MDC) and advanced filtering techniques such as Independent Component Analysis (ICA) as proposed by Kosmyna et al. [2015a] and Kosmyna et al. [2015b] for MI. The architecture allows for an incremental training and single trial classification in order to minimize training time. A short training time is an important element in making BCIs better usable for everyday applications. We first present the signal processing and acquisition aspects, followed by distance measures and the principles behind the MDC approach. We will discuss the rationale behind the choice of single trial classification later on, in the experimental protocol description (Section 6).

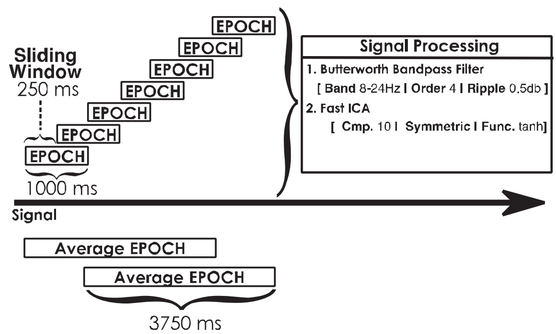


Fig. 3. Signal epoch segmentation and averaging in our system.

5.1. Signal Processing & Acquisition

Figure 3 gives a graphical representation of how signals are filtered and divided into 1s epochs that each overlap over 250ms. The average of the 250ms window is subtracted from each of the epochs. The following filtering steps are performed on each 1s epoch prior to the averaging:

- The Butterworth filter allows selecting the appropriate frequencies for the modality and discarding unwanted frequencies. The frequencies filtered by the Butterworth filter are 8Hz–30Hz on pre-epoch signals for Conceptual Imagery, as reported by Simanova et al. [2010].
- The Fast ICA [Hyvärinen and Oja 1997] algorithm projects the signal data in a space in which data points are maximally independent, essentially separating task related sources from noise sources and other interference. We computed 10 components (determined empirically). We used a GPL Java implementation.
- Then, we produce average epochs that allow us to smooth the signal and remove some of the variability. Each average epoch is formed from the average of overlapping within-epoch signals of five consecutive epochs. If the sliding window is 250ms, then the average epoch formed from five epochs will be 3750ms long.

The system thus produces an average epoch four times per second. Each epoch is then used for feature extraction and for classification. Thus, the classifier will yield one classification per second. Given that ICA is rather costly to compute, anything less than one second led to sub real-time performance on the machine we performed the processing on (2012 MacBook Air, i7@2.9GHz).

Our classifier only requires minimal training data to start functioning, as our aim was to reduce that training time to a single calibration trial and per class. This allowed us to capture a reference signal for each of the classes. In the case of our system, we captured full averaged epochs as references. Thus, the calibration for each class lasted 3750ms and we left 2s of rest between each calibration trial so that users could unwind briefly and prepare for the next trial. Figure 4 graphically shows the process by which we capture the reference signal for a particular BCI task. A single average epoch is used as class reference; this is why the calibration phase is synchronized with the beginning of an average epoch.

5.2. Feature Extraction & Classification

As described in Figure 4, for each classification, we take the current average epoch and compute the distance between this current epoch and each of the reference signals for each of the classes. We then select the class that has the lowest distance value (arg min as shown in Figure 5). If there are several training trials and thus class references,

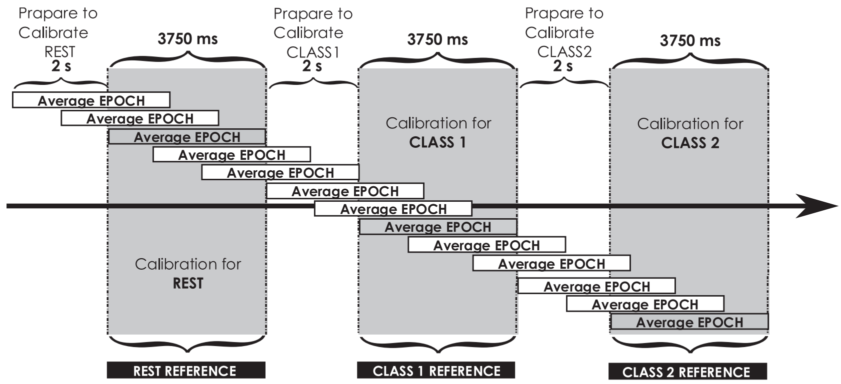


Fig. 4. Detailed view of the calibration phase.

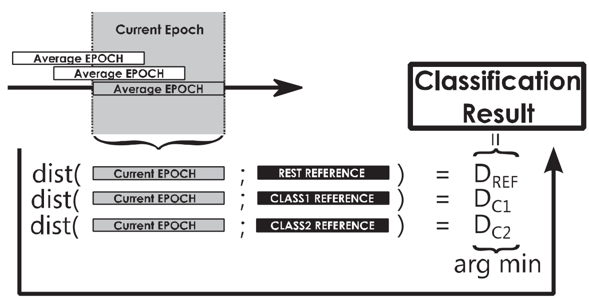


Fig. 5. Our classifier with three classes and one reference epoch per class, using a multinomial distance (e.g. Mahalanobis).

there are several classification decisions (one per class reference). We combine the decisions through a majority vote.

6. EXPERIMENTAL PROTOCOL

In order to evaluate the potential of the proposed approach, we first draw a parallel with comparable methods and situate the proposed approach as a natural extension in order to adapt the existing work to training BCI systems based on machine learning. More specifically, Flatla et al. [2011] have implemented and proposed the implicit calibration of an input modality within a game by using specific game mechanics for different types of calibrations (**In**-**G**ame **T**raining, **i**mplicit = IGTi+), however the game just serves for the calibration and the application task is unrelated (**T**ask is **not** the **G**ame = TG−). On the other extreme, Nacke et al. [2011] have used BCIs and other physiological signals in a game in which the task is to adapt and control the game (the **T**ask **is** the **G**ame = TG+), but **t**raining is not performed in-game (**no** **I**n-**G**ame **T**raining = IGT−). To our knowledge, the combination of in-game training in which the task is to play the game (TG+, IGT+) has never been explored. Moreover for in-game training (IGT+), two subfactors can be identified. Training can be encoded in the game implicitly as game mechanic as stated earlier. Alternatively, **t**raining can also be encoded in the game environment as an explicit **s**ynchronous and obligatory step that gives players explicit instruction (**In**-**G**ame **T**raining **s**ynchronous = IGTs+).

Table I shows the confusion table of all the possible conditions and highlights what conditions have already been studied in the literature and what conditions haven't but that are of relevance for the present study (training integrated in-game + applying the

Table I. Condition Combinations Depending on Whether the Training was in Game (Explicit Synchronous ITGs+; Implicit ITGi+) or not (ITG−) and Whether the Task in Which the BCI is Applied is the Same Game (TG+) or not (TG−). The White Color Indicates That it has Been Attempted in the State of the Art, but That it's not Covered in this Paper. Blue Means That it's Been Attempted in the State of the Art and That we use it in This Work as a Baseline. Green means it's not Been Attempted in the State of the Art and That it is Part of our Experiment

•	IGT		
TG	–	s+	i+
+	Standard BCI for control & adaptation.	Direct integration of standard BCI training in a game that we want to control with the BCI. Never attempted. This requires an extension of the work by Flatla et al. [2011] to Machine learning based active and reactive BCIs and its application for the control of the same game.	Seamless integration of BCI training through **priming** in a game in order to control that same game. Never attempted when applied to control the same game.
–	Standard BCIs applied to other control tasks.	Direct integration of standard BCI training in a game in which the BCI is meant for different control tasks.	The work of Flatla et al. [2011] propose a seamless calibration framework that is applicable to passive and some reactive BCIs, but not for active machine learning based BCIs.

BCI for the control of the game in the continuity of training). The first step towards that goal is to achieve a direct and explicit integration in the standard BCI synchronous training and then to find a game mechanic that masks the synchronous training process through the proposed use of priming and thus to achieve an extension of Flatla et al. [2011] for training BCIs based on machine learning.

Consequently, we want to compare whether or not the explicit integration of synchronous BCI training improves the game experience and immersion and then to verify whether masking the synchronous training in a game mechanic through the use of priming leads to an improvement over the inclusion of explicit training with relation to an out-of-game synchronous BCI training baseline. In other words, we are only interested in the three conditions: (IGT−, IGTs+, IGTi+).

In order to qualify what "improvement is" we designed a questionnaire based on the core GEQ [IJsselsteijn et al. 2008] that measures five indicators. For each indicator, there are six questions on a Likert scale between 0 and 5 for immersion and between 0 and 4 for the other indicators. See Appendix B in Nacke [2009] for the complete questionnaire and a complete description of how the scores are calculated. Each indicator is expressed as the arithmetic mean of the answer to each of the questions.

The five indicators are as follows:

- Immersion describes the degree to which a player is engaged in the game playing experience while retaining awareness of their surroundings [Nacke and Lindley 2008].
- Flow is the feeling of enjoyment that results from a balance between challenge and skill [Nacke and Lindley 2008].
- Competence is the feeling in players that they are skillful or successful [Hoogen and Kort 2008].
- Tension as the name indicates corresponds to a state of tension of the player who is constantly anticipating what will happen next [Nacke and Lindley 2008].
- Positive affect occurs when the gameplay experience elicits a positive emotion or impression on the user [Nacke and Lindley 2008].

- Negative affect occurs when the gameplay experience elicits a negative emotion or impression on the user [Nacke and Lindley 2008].

In this study we took into account only Immersion, Flow, and Competence (Cronbach's alpha was respectively 0.79, 0.83, 0.81) as Tension, Positive, and Negative Affect were not very reliable (Cronbach's alpha was respectively 0.64, 0.49, 0.52).

Moreover, we also evaluate BCI performance by selecting six dark passages in the level in which a flashlight is required, and six enemies that must be killed to progress. We count the number of errors before a correct classification is produced.

Based on the previous conditions, we make the following hypotheses that we need to verify experimentally:

- H.1). The integration of synchronous BCI training explicitly in the game (IGTs+) leads to significantly higher average values for the immersion (H.1.i), flow (H.1.f), and competence (H.1.cmp.) factors compared to the baseline (IGT−).
- H.2). The masking of the synchronous in game BCI training through priming leads (IGTi+) to higher average values for the immersion (H.2.i), flow (H.2.f), and competence (H.2.cmp.) factors compared to the baseline (IGT−) and to explicitly integrated synchronous training (IGTs+).

6.1. Subjects

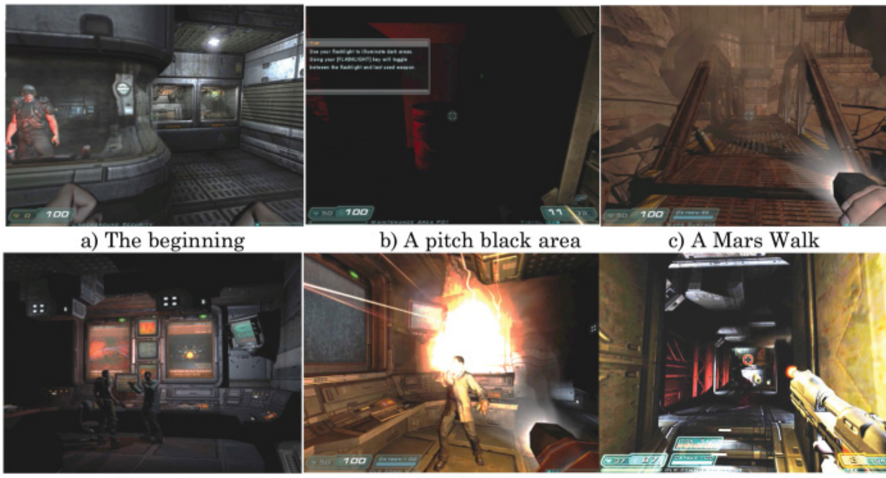
For the experimental evaluation, we need three groups of subjects to match the three conditions. To that effect, we selected 36 subjects aged between 23 and 34 years old in good physical and mental health, 18 male and 18 female. The subjects were selected among passers-by on a large university campus. We screened the users on gaming experience, good health and no prior BCI experience. We selected all the participants to have moderate gaming experience (casual gamers only) who had never played Doom 3 but who had FPS experience, so that varying familiarity with the game would not bias the experiment. Each of the experimental groups (IGT−, IGTs+, IGTi+), were formed randomly by 12 people from total population. We follow the same distribution as the entire experimental population so as to keep a representative subsample. IGT− had 12 subjects (6 male, 6 female) between 24 and 32 with a median age of 28. IGTs+ had 12 subjects (6 male, 6 female) between 23 and 30 with a median age of 26. IGTi+ had 12 subjects (6 male, 6 female) between 25 and 34 with a median age of 29.

6.2. Game Modifications

For the integration in Doom 3, we modified the event handling code of the engine by adding a client/server communication with our BCI system through the Open Sound Control protocol. We intercept game events and send the name of the triggering entity to the BCI. Similarly, we map BCI classification outcomes to in-game events and send the event to the game for actuation. With this modification in place, we had to modify game levels or craft new levels by adding triggers with names known to the BCI system in order to trigger the synchronous training trials in the BCI and to notify the game that the training trial is complete.

Given that the context of the paper is the immersive integration of BCIs with games, we chose to base the experiment on the second level of the game, in which the user is first exposed with the gameplay mechanics.

This makes sense as the game already includes instructions for the core mechanics specifically in this level. Namely, the user is introduced to the first weapon of the game, a pistol, as well as the flashlight. Doom 3 is a very dark game and the code mechanic of the game is to switch to the flashlight to canvas the levels and detect the appropriate path to follow and when enemies appear, to switch to an appropriate weapon to deal with the threat. For this reason and given the nature of conceptual imagery and the fact



a) The beginning b) A pitch black area c) A Mars Walk

d) The cinematic e) The first enemy f) It's time

Fig. 6. The level used in the experiment, "Mars City Underground".

that it is suitable to detect imagined conceptual representations, we chose to build the BCI to detect the conceptual representations for "flashlight" and for "weapon" in order to map them to the *toggle flashlight* action and to the *select gun* action. The activations in the brain for concepts that are not related to motor activity do manifest in the motor cortex, allow users to move in a limited fashion. Thus, the character movement was controlled through the keyboard and we disabled keyboard controls for the flashlight and for weapon selection in order to prevent the users from cheating by not using the BCI.

The game level itself consists in a rescue mission that takes place on Mars. The character is a marine recently arrived in Mars city (a futuristic colony on the red planet) and is assigned the mission to rescue and retrieve a scientist fled to an old communications building. The player first has to take the required gear (armor, gun, clips), a guard gives instructions and directions about the mission and opens the door to a underground area that has to be crossed to arrive to the scientist.

Shortly, thereafter the player crosses a pitch back room and is told to use the flashlight. Right after that, the player crosses a small portion of the Mars surface to reach the abandoned building. On entering, a cinematic is triggered that shows cataclysmic events and that results in the scientist turning into a zombie and attacking. This is the first moment in which the weapon is used (See progression in Figure 6).

For the IGTi+ condition, the player starts at the beginning of the level and the priming for the BCI training is performed when the player first needs to use the flashlight and then when the payer must use the gun for the first time. For IGT− and IGTs+, the BCI training is performed either before the game session altogether, or in a separate section of the map meant for BCI training. After their training sessions, both for IGT− and IGTs+ the player starts at the beginning of the level played by IGTi+.

6.3. Training Protocols

First, let us clarify the vocabulary pertaining to BCI training. BCI training is composed of several so-called training "trials" for each of the classes. A training trial is a period of time repeated one or more times for each BCI class and is composed of the following three phases:

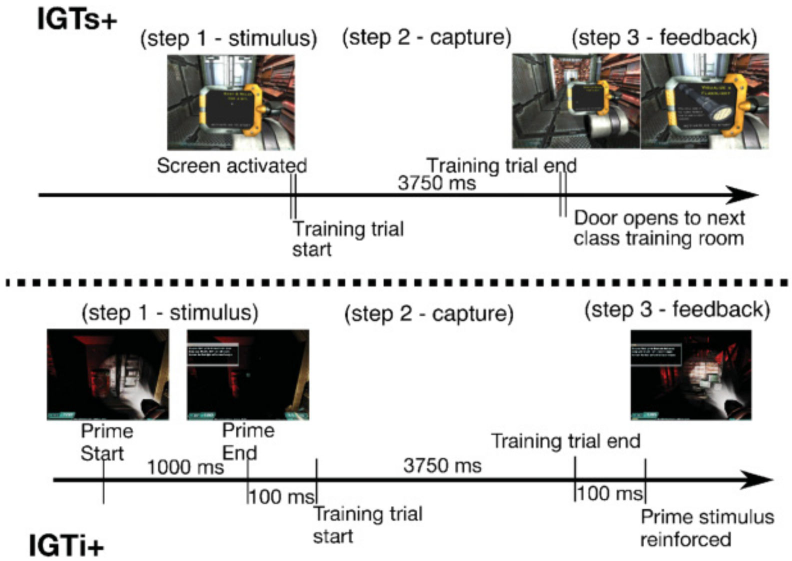


Fig. 7. The IGTs+ versus IGTi+ training protocols for each class.

1. A *stimulus* is presented to the user that gives an indication of the BCI class involved in the current trial. The *stimulus*, in BCIs and in particular with the present conceptual BCI, indicates the user what concept to imagine and/or visualize.
2. After the *stimulus*, EEG signals from the user are *captured* for a fixed duration (here for 3750ms).
3. *Feedback* indicates that the capture is complete. In a standard synchronous training, the end of the capture is marked by the beginning of a resting phase and is typically indicated by a message. In our synchronous in-game training (IGTs+) we let the user move on to the next room for the next trial. In IGTi+, we trigger the action that will be associated to the class that corresponds to the current trial.

When enough trials have been performed for each of the classes (typically 10–20 per class with a standard supervised system, here we use only one), the classifier used in the BCI system is trained with the captured signals associated to the expected BCI class. All the training protocols described later correspond to steps 1 to 3 above.

Let us now describe and illustrate the three training protocols (see protocol timelines in Figure 7). For the BCI, the three classes to train (due to the synchronous nature of the system) are “Flashlight”, “Weapon” and “Other activity/Resting state”.

- IGT−: We perform the training of the BCI outside of the game using a training interface that represents each of the classes by an image that serves as the semantic stimulus. A flashlight represents the “flashlight” class, a weapon represents the “weapon” class and a blank square represents the resting state (Figure 8). We capture one trial per class. The images representing the classes are magnified to signify to the player that they must imagine the concept associated to that class (step 1 from above – *stimulus*). Then follows a capture period of 3750ms (step 2 – *capture*), after the trial there were 5s of rest, after which a message prompted the player to get ready for the next class in 2s (step 3 – *feedback*). After the BCI is trained, the player starts playing at the beginning of the level of the game.
- IGTs+: We create a separate set of rooms in the level for BCI training that lead to the moment after the first cinematic. In the training rooms, for each BCI class the user

Fig. 8. Images used during the synchronous in-game training to represent the three BCI classes, “flashlight”, “weapon”, and resting state.

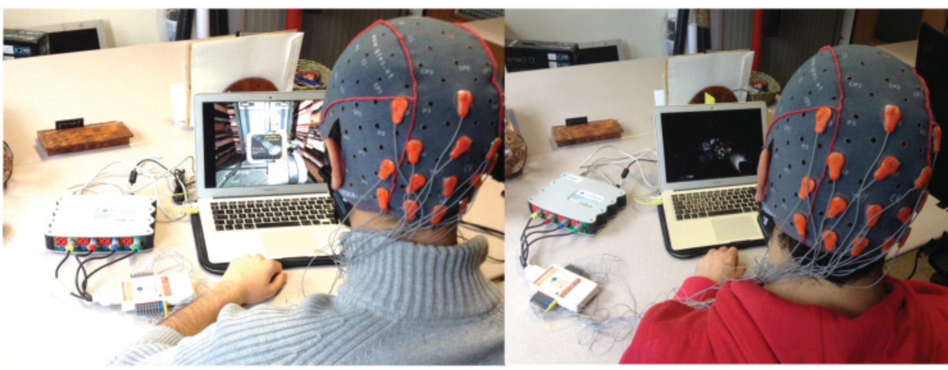


Fig. 9. Explicit training for the flashlight action (left) and implicit training for the flashlight (right).

Fig. 10. The flashlight in action.

arrives in front of a screen with a GUI with instructions asking the user to imagine a flashlight, a weapon or to stay at rest (step 1 – *stimulus*). When the in-game GUI is activated, a BCI capture period of 3750ms for the class in question begins (step 2 – *capture*). When the trial is finished, a door opens and leads to a similar room for the training of the next class (step 3 – *feedback*). See the protocol timeline in Figure 7 and an illustration of a player undergoing the explicit integrated training of the flashlight in Figure 9 (left). Figure 10 illustrates the flashlight in action.

- IGTi+: The player starts from the beginning of the level. Before players retrieve their gear, movement is impossible and that is in which we capture the signals for the “Rest” class (step 2 – *capture*, step 1 and step 3 are not necessary for the rest class as it corresponds to any activity other than the imagination of flashlight and weapon. The dark passage in the level with the flashlight tip is used for the moment in which priming is performed for the flashlight (step 1 – *stimulus*, lasts 1s. The prime exposure time is a compromise, so that the time of priming exposure matches the duration of one epoch of the BCI training trial. The typical time in the literature is of the order of 400ms), followed by the 3750ms capture period for "flashlight" (step 2 – *capture*). Then, the flashlight is activated (step 3 – *feedback*). Later, the moment the first enemy appears is used for the priming (step 1 – *stimulus*) and subsequent capture (step 2 – *capture*) of the signals for the BCI class "weapon". It is likely that fear/panic are also captured during training, which would lead to incorrect classifications if the user wishes to select the weapon prior to engagement. Another moment in which the weapon could be trained is the first time the weapon is introduced at the beginning of the level, which could overcome the issues of panic/fear. Given that we wanted the prime to be the weapon firing, the second situation was deemed less adequate. Figure 9 (right) shows a player during the implicit flashlight prime training. Figure 10 illustrates the flashlight in action.

A possible confound is that on a per-trial basis, IGTi+ (according to Figure 7) gets 1200ms more time per-trial due to the priming durations (which may indeed constitute a bias) as opposed to IGTs+ training. However, while for IGTi+, the training is integrated in the full game level (unnoticeable and designed in such a way as to not make the completion of the level longer by blending in the narrative), for IGTs+ training takes place in a separate level preceding the game level. Therefore, IGTs+ technically gets the most game time. Although, arguably the in-game training level is more similar to the external interface of IGT−, not to mention that the actual gameplay elements controlled by the BCI are not practiced at all in the training level. Moreover, the gameplay segments are identical from the perspective of the player for IGTi+ and IGTs+/IGT−. Finally, competence was not a factor considered for the experiment and thus the additional game time does not influence the results in any meaningful way. We argue that in the end no condition is favored over another within the scope of our evaluation objective.

6.4. Hardware

We used the g.tec USBAmp, a high-end bio signal (Electroencephalography or EEG, Electrocardiography or EKG, others) amplifier. We use it coupled with 16 electrodes on a standard 10–20 EEG cap. We use a sampling frequency of 512.

For the electrode placement of conceptual imagery, we mostly followed the placement proposed by one of the pilot studies on this BCI modality by Simanova et al. [2010], in which 64 electrodes are positioned in an equidistant manner so as to locate and study the underlying phenomena.

They make three experiments in which visual, textual, and auditory stimulation cues are used and find that visual cues lead to the best performance. Moreover, they study the activation over all the 64-electrode mesh and identify the zones with the most activation to be towards the center of the visual, and occipito-parietal cortices. As a consequence, we made the choice of using visual cues for our implementation as they yield the best activation. Moreover, we placed our electrodes on the standard 10–20 electrode placement system so as to cover the activation area reported by Simanova et al. Figure 11 shows the electrode positions.

6.5. Offline Validation of the BCI System

Prior to the experiments, we validated the performance of the BCI in an offline synchronous setting independent of any particular task. The offline validation of a BCI system is standard practice for BCI research and is important to gauge the validity

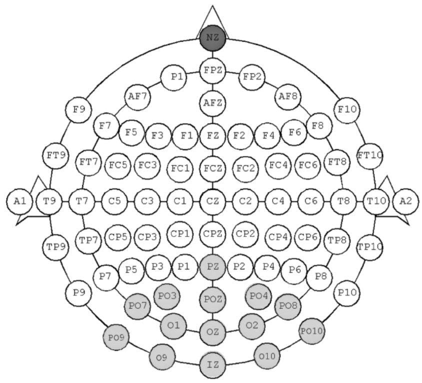


Fig. 11. The electrode placement for conceptual imagery. The dark gray electrode is the reference.

of the signal processing and classification approaches used. The evaluation consisted in capturing a large number of training trials for each class and of selecting a first subset treated as if we did not know what class they belonged to (test set) and then of selecting a second subset in which class labels are kept (training set). We create 10 such partitions and use one partition for training and 9 for testing, which allows us to make sure the model of the classifier generalizes well to unseen instance (10 times 10-fold cross-validation). We evaluated the three distance measures that can be used with our system in order to see what the best distance measure (if any) is for our current Conceptual Imagery setting. The distances considered are the following:

- The Mahalanobis distance is a multivariate measure that take correlation into account and that measures the number of standard deviations (in an elliptical shape) of distance between an instance and a distribution of points. The Mahalanobis distance is computed on the covariance matrix of two signals. It is relatively robust as it is scale invariant. If the matrix has unit variance and standard deviation, it is equivalent to the generalized Euclidian distance computed on the signals [Lotte et al. 2007].
- The Riemannian distance takes a step further, and assumes the distribution has a non-Euclidean topology. It first projects each point of the distribution on a tangent Euclidean plane that conserves relative distances. It has been used for several BCI paradigms including P300, SSVEP, and MI. For P300 and SSVEP covariance matrices are crafted specifically, but for MI the covariance matrix is used directly. Here, we use the covariance matrix directly [Congedo et al. 2013].
- The Spearman Rank Correlation Distance is a nonparametric measure of correlation inverted and scaled to become a distance. The Spearman rank correlation captures linear relationships between signals in terms of their relative behaviors and does not capture time shifts. One must rely on averaging to get rid of alignment issues.

The measures have been applied the most to BCIs for distance based-classification and obtained good results, which is why we have selected them [Lotte et al. 2007]. There are no parameters to set for any of these measures.

We selected 10 subjects from ages 25–32 in good health and with no prior BCI experience that did not take part in any of the experiments reported in Section 7. The analysis was done over the signals of the 10 subjects captured over the course of 40 training trials per subject and per class (120 trials per subject) following the acquisition protocol detailed in Section 5 for the CLBCI system. The CLBCI system is usable with three different distance measures, thus we trained and evaluated the

Table II. 10-fold Cross-Validated Median Accuracy of our BCI System on 10 Subjects over 10 Sessions for Conceptual Imagery (99% Confidence Intervals)

Distance Measure	Median Classification Accuracy	99% Confidence Interval
Maha.	71.32%	2.93%
Riem.	73.14%	2.73%
Corr.	72.23%	2.81%

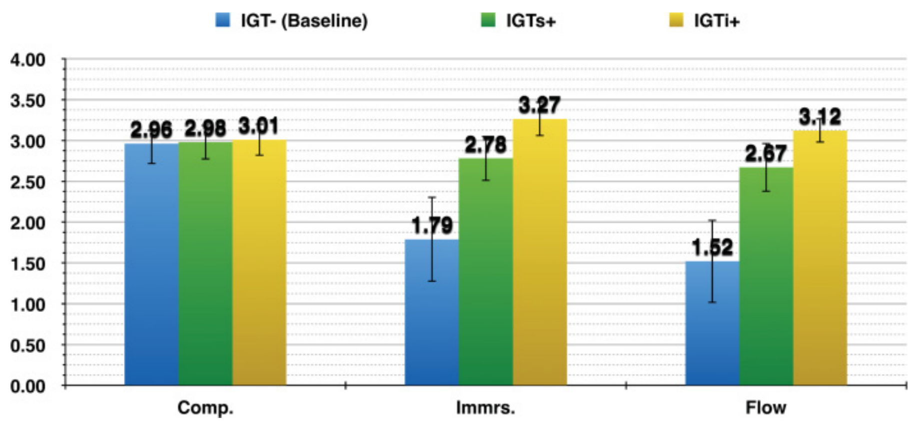


Fig. 12. Median results for the questionnaire for all three conditions (IGT−, IGTs+, IGTi+) for the Competence, Immersion and Flow indicators. Error bars are 95% CDF-based nonparametric confidence intervals.

average BCI classification accuracy for each measure across the 10 subjects with 10-fold cross-validation.

We can now look (Table II) at the cross-validation results from the off-line classifier evaluation to validate the BCI system and to estimate the best distance measure to use in the current setting. We did not find any significant difference in performance for any one of the measures. The results are significantly different from the results that would be obtained from random classification: Müller-Putz and Scherer [2008] estimate the upper-bound of the empirical random classification accuracy at 65% with a confidence interval of 0.1% for 40 trials per class.

Although no distance measure is conclusively better, we chose to use Mahalanobis distance for the experiments as it had the greatest absolute median classification accuracy and the lowest standard deviation among the three.

7. RESULTS

7.1. Analysis

Given the low number of subjects for each condition, we used the Shapiro–Wilk test and found a *p-value* of $p = 0.04439$, which means there is insufficient evidence to accept that the null hypothesis of the normality of the distribution of the data is valid. Consequently, we use the Kruskall–Wallis test to measure for significant group effects and then use the Mann–Whitney U-test for post-hoc pairwise analysis with a False Discovery Rate (FDR) *p-value* adjustment for multiple comparisons.

For each condition we averaged the scores of the questions pertaining to each factor, and reported the results in Figure 12 as a histogram. The error bars represent the 95% cumulative distribution function (CDF)-based nonparametric confidence intervals.

For competence, there are no significant group differences (Kruskall–Wallis, $H = 3.12$, d.f. $= 2$ $p = 0.18769$). This means that the perceived performance of the BCI did not

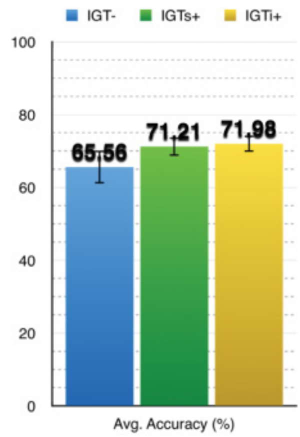


Fig. 13. Median classification accuracy (%) for BL, IGTs+ and IGTi+. Error bars are 99% CDF-based nonparametric confidence intervals.

change significantly whether the training was outside of the game, explicit synchronous or implicit. This result invalidates the hypotheses (H.1.cmp.) and (H.2.cmp.).

For immersion and flow, Kruskall–Wallis found significant rank differences between the groups, (KW, H = 6.33, d.f. = 2, p = 0.03265) and (KW, H = 7.21, d.f. = 2, p = 0.02598). Similarly, still for flow and immersion, all the pairwise comparisons are significant according to the Mann–Whitney U-test.

For reference, in terms of the difference between IGTs+ and IGTi+ for flow (the smallest effect) M–W.U yields U = 48 > 37 (critical value for p = 0.05). For both immersion and flow IGT− achieves a low score (1.79, 1.52). The score almost doubles for ITGs+ (2.78, 2.67) and gains one third more for IGTi+ (3.27, 3.12). Thus, we validate hypotheses H.1.i, H.2.i, H.1.f, H.2.f.

We also evaluated online BCI classification accuracy across the three conditions, by considering six dark passages and six enemies (as described in Section 6) and observing for each of them if the classification result matched the desired outcome (selection and keeping the flashlight up for dark passages and selection and keeping the weapon up for enemies). Figure 13 shows the median values and confidence intervals for BCI performance in terms of classification accuracy and classification time. For classification accuracy, there is a significant group effect (KW, H = 30.494, d.f. = 2, p = 0.00000[1]). There is no significant difference in the mean between IGTs+ and IGTi+ (MWU, U = 74, Z = –1.5762, Critical U (0.05) = 72, p = 0.05705). However, both IGTs+ and IGTi+ are a significant improvement over BL (MWU, U = 0, Z = –4.6455, Critical U (0.05) = 72, p = 0.00000) and (MWU, U = 0, Z = –4.6455).

7.2. Opinions of Users

After the experimental sessions, we asked our players some informal questions about their experience. The subjective opinions of our players were overly positive for the integrated training and especially for the fully integrated training that felt the most natural to them. They found the BCI to be a motivating modality as they were excited to discover what a BCI can do. For the synchronous integrated training players were actually forced to imagine the flashlight within the game environment and the players were more positive about the training, although it felt artificial to some of them to have to activate a GUI each time and to wait for the game to begin. Overall, the implicit

[1]The p-value is not zero, however when rounded to the five digits we report, the result is zero.

integrated training garnered the most positive remarks, as some users didn't even realize they were training a BCI (except for the fact that they had a cap strapped to their heads). One of the users commented that "the experiment was really cool and interesting, if only the accuracy was a bit higher and if there were less cables, I'd really consider using something like this to play/replay other games such as Half Life".

8. DISCUSSION AND LIMITATIONS. FUTURE WORK

We expected that better immersion during training would lead to better BCI performance. Although the actual BCI performance increased, the perceived competence did not change. This suggests that better immersion influences BCI performance on a subconscious level.

In the implicit training phase, we were limited to one trial as we wanted to use passages from a real level only that matched the introduction of the concepts in the game. This limitation was not present in the explicit external and internal trainings as we could have as many trials as needed to reach the best possible BCI performance. However, we use one trial everywhere to be able to compare BCI accuracy between the groups fairly. Single trial BCIs typically achieve performance around 60%, which explains the average level of competence felt by users. It would be interesting to integrate more training trials in the game environment (by slightly modifying the level) and to study the relationship between the number of training trials and the degree of immersion. One can hypothesize the adding more training trails will increase the perceived competence and BCI accuracy and decrease the flow and immersion, unless trials are very well camouflaged. An inquiry about general techniques and guidelines that help design and modify game levels to integrate more training trials is very pertinent and required.

As expected, the standard BCI in which training is performed outside of the game, achieves a low degree of immersion and flow as opposed to the explicit and implicit training integration. This suggests that the principal obstacle to flow and immersion when BCIs are applied to games is that the BCI training is separate from the game. When the training is made implicit and integrated in the gameplay, the flow and immersion reach good levels. This suggests that the implicit integration through priming is natural to the users and did not seem out of place. The alternative could have been that suddenly losing the flashlight or weapon creates a panic in the player and interferes with the BCI classification process.

The main limitations are that we need to find a way of integrating more training trials seamlessly in order to obtain a better sense of competence in users and to allow for a better BCI control. Another limitation is the determination of the duration of the priming and the duration of the BCI training trails. Longer training trails and priming times may lead to better performance, although the only way of finding this out is to perform an empirical grid search over training and priming times to determine which combination is the best.

More generally, we considered only regular video game players. This means that our conclusions cannot be readily transposed to nongamer populations, in which the importance of immersion could be lower.

Despite these limitations the experiment gives interesting insights into training protocol for BCIs, when reaching a situation in which users do not realize they are training a BCI is an important step towards out-of-the-lab uses of BCIs. Moreover, even for experienced gamers, the BCI dimension is potentially an added gameplay experience.

The hypotheses we state must be validated in larger scale longitudinal studies to generalize them for long-term use and for other application fields than BCIs, in which training is implicitly integrated in the environment.

In theory, this training strategy can be generalized to other machine learning processes that rely on the reactions of users (beyond EEG). However, they're no such applications to date. We theorize that this system could be used to train an AI system for nonplayer antagonists that learn from the reaction of players to particular priming stimuli.

9. CONCLUSIONS

In this article, we proposed a novel implicit integration of BCI training protocols in video games. We first proposed a protocol in which the synchronous BCI training protocol is directly transposed in the game environment. Then, we improve on the integration by masking the steps of the training in elements of the gameplay and by using semantic and visual priming to elicit the expected player EEG response required for training the BCI system. We conduct experiments on 36 subjects with video game experience and no BCI experience in order to implement the two protocols and train the BCI over a single trial. We compare the two protocols to a standard BCI training outside of the game. We evaluate the best training protocol using the core GEQ associated with a predictive reliability analysis (Cronbach's alpha). We find that flow, immersion and competence are the only indicators that are reliable (Cronbach's alpha above 0.7). A single training trial leads to limited performance (a feeling of competence of around three out of four for all conditions), the increase in flow and immersion by integrating BCI training explicitly is sizable (2.78/4 and 2.67/4 instead of 1.79/4 and 1.52/4). The further implicit integration of BCI training led to a more sizable increase (3.27/4 and 3.12/4). However we suspect that if more training trials are performed, the immersion associated with the implicit training protocol may diminish. The proposed protocol has a potential towards the design of better BCI training protocols and more generally integrate machine learning processes that require user intervention in video games (e.g. training a nonplayer character AI that adapts to subtle reactions of the player triggered by priming). More complete studies are required in other applications than video games (e.g. VR) and over longer usage periods.

REFERENCES

Alvaro Barbero and Moritz Grosse-Wentrup. 2010. Biased feedback in brain-computer interfaces. *J. Neuroeng. Rehabil.* 7, 1 (January 2010), 34 pages.

Leonard Berkowitz and Anthony Lepage. 1967. Weapons as aggression-eliciting Stimuli. *J. Pers. Soc. Psychol.* 7, 2 (1967), 202–207.

Irving Biederman and Eric E. Cooper. 1992. Size invariance in visual object priming. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 1 (1992), 121–133.

Benjamin Blankertz, Florian Losch, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Klaus Robert Müller. 2008. The Berlin brain - computer interface: Accurate performance from first-session in BCI-naïve subjects. *IEEE Trans. Biomed. Eng.* 55, 10 (Oct. 2008), 2452–2462.

Danny Plass-oude Bos, Michel Obbink, Anton Nijholt, Gido Hakvoort, and M. Christian. 2010. Towards multiplayer BCI games. In *BioS-Play: Workshop on Multiuser and Social Biosignal Adaptive Games and Playful Applications.* 1–4. http://www.physiologicalcomputing.net/?page_id=1767.

Pierre Chalfoun and Claude Frasson. 2012. Cognitive priming: Assessing the use of non-conscious perception to enhance learner's reasoning ability. In *Intelligent Tutoring Systems*, Stefano A. Cerri, William J. Clancey, Giorgos Papadourakis, and Kitty Panourgia (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 84–89. http://doi.org/10.1007/978-3-642-30950-2_12.

Marco Congedo et al. 2011. "Brain Invaders": A prototype of an open-source P300-based video game working with the OpenViBE platform. In *5th International BCI Conference*. Graz, Austria, 280–283.

Marco Congedo, Alexandre Barachant, and Anton Andreev. 2013. A new generation of brain-computer interface based on riemannian geometry. CoRR abs/1310.8. Retrieved from http://arxiv.org/abs/1310.8115.

Tamara Denning, Kevin Bowers, Marten van Dijk, and Ari Juels. 2011. Exploring implicit memory for painless password recovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, USA, 2615–2618. DOI:10.1145/1978942.1979323

Jerry Alan Fails and Dan R. Olsen Jr. 2003. Interactive machine learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI'03)*. ACM, New York, NY, USA, 39–45.

Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. Human model evaluation in interactive supervised learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, USA, 147–156.

David R. Flatla, Carl Gutwin, Lennart E. Nacke, Scott Bateman, and Regan L. Mandryk. 2011. Calibration games: Making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST'11)*. ACM, New York, NY, USA, 403–412. DOI:10.1145/2047196.2047248

Marcel van Gerven, Jason Farquhar, and Peter Desain. 2013. Detecting semantic priming at the single-trial level. *PLoS One* 8, 4 (2013), 1–7.

Daniel Graybill, Janis R. Kirsch, and Edward D. Esselman. 1985. Effects of playing violent versus nonviolent video games on the aggressive ideation of aggressive and non aggressive children. *Child Study J.* 15, 3 (1985), 199–205.

C. Shawn Green and Daphne Bavelier. 2003. Action video game modifies visual selective attention. *Nature* 423, 6939, 534–537. Retrieved from http://www.nature.com/nature/journal/v423/n6939/suppinfo/nature01647_S1.html.

C. Shawn Green and Daphne Bavelier. 2007. Action-video-game experience alters the spatial resolution of vision. *Psychol. Sci.* 18 (2007), 88–94.

C. Shawn Green and Daphne Bavelier. 2006a. Effect of action video games on the spatial distribution of visiospatial attention. *J. Exp. Psychol. Hum. Percept. Perform.* 32 (2006), 1465–1478.

C. Shawn Green and Daphne Bavelier. 2006b. Enumeration versus multiple object tracking: The case of action. *Cognition* 101 (2006), 217–245.

Anthony G. Greenwald, Sean C. Draine, and Richard L. Abrams. 1996. Three cognitive markers of unconscious semantic activation. *Science* 273, 5282, 1699–1702. DOI:10.1126/science.273.5282.1699

Lane Harrison, Drew Skau, Steven Franconeri, Aidong Lu, and Remco Chang. 2013. Influencing visual judgment through affective priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, New York, NY, USA, 2949–2958. DOI:10.1145/2470654.2481410

Wouter Van Den Hoogen and Yvonne De Kort. 2008. Exploring behavioral expressions of player experience in digital games. In *Proceedings of the Workshop for Facial and Bodily Expressions for Control and Adaptation of Games*. CTIT, Enschede, The Netherlands, 11–19.

Aapo Hyvärinen and Erkki Oja. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9 (1997), 1483–1492.

W. IJsselsteijn, W. van der Hoogen, C. Klimmt, Y. de Kort, C. Lindley, K. Mathiak, K. Poels, N. Ravaja, M. Turpeinen, and P. Vorderer. 2008. Measuring the experience of digital game enjoyment. In *Proceedings of Measuring Behavior*, A. J. Spink, M. R. Ballintijn, N. D. Bogers, F. Grieco, L.W.S. Loijens, L. P. J. J. Noldus, G. Smit, and P. H. Zimmerman (Eds.). Maastricht, The Netherland, 26–29.

Inaki Iturrate, Luis Montesano, and Javier Minguez. 2010. Single trial recognition of error-related potentials during observation of robot operation. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'10)*. 4181–4184. DOI:10.1109/IEMBS.2010.5627380

Ashish Kapoor, Bongshin Lee, Desney Tan, and Eric Horvitz. 2010. Interactive optimization for steering machine classification. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'10*. ACM Press, New York, New York, USA, 1343–1352.

Simon P. Kelly, Edmund C. Lalor, Richard B. Reilly, Senior Member, and John J. Foxe. 2005. Visual spatial attention tracking using high-density SSVEP data for independent brain – computer communication. *IEEE Trans. neural Syst. Rehabil. Eng.* 13, 2 (2005), 172–178.

Nataliya Kos'myna, Franck Tarpin-Bernard, and Bertrand Rivet. 2013. Towards a general architecture for a co-learning of brain computer interfaces. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 1054–1057. DOI:10.1109/NER.2013.6696118

Nataliya Kosmyna and Franck Tarpin-Bernard. 2013. Evaluation and comparison of a multimodal combination of BCI paradigms and eye tracking with affordable consumer-grade hardware in a gaming context. *IEEE Trans. Comput. Intell. AI Games* 5, 2 (2013), 150–154.

Nataliya Kosmyna, Franck Tarpin-Bernard, and Bertrand Rivet. 2015a. Adding human learning in brain computer interfaces (BCIs): Towards a practical control modality. *ACM Trans. Comput.-Hum. Interact.* 22, 3, Article 12 (May 2015), 37 pages. DOI:10.1145/2723162

Nataliya Kosmyna, Franck Tarpin-Bernard, and Bertrand Rivet. 2015b. Towards brain computer interfaces for recreational activities: Piloting a drone. In *Proceedings of INTERACT 2015*. Bamberg, Germany.

Marta Kutas and Kara D. Federneier. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 12 (2000), 463–470.

Marta Kutas and S. A. Hillard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 207, 4427, 203–205. DOI:10.1126/science.7350657

Karl LaFleur, Kaitlin Cassady, Alexander Doud, Kaleb Shades, Eitan Rogin, and Bin He. 2013. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface. *J. Neural Eng.* 10, 4 (2013), 046003.

Anatole Lécuyer, Fabien Lotte, RB Reilly, and R. Leeb. 2008. Brain-computer interfaces, virtual reality, and videogames. *Computer (Long. Beach. Calif).* 42, 10 (2008), 66–72.

Sheena Lewis, Mira Dontcheva, and Elizabeth Gerber. 2011. Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, USA, 735–744. DOI:10.1145/1978942.1979048

Renjie Li, Uri Polat, Walter Makous, and Daphne Bavelier. 2009. Enhancing the contrast sensitivity function through action video game training. *Nat. Neurosci.* 12 (2009), 549–551.

F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, 2 (June 2007), R1–R13. DOI:http://dx.doi.org/10.1088/1741-2560/4/2/R01

Fabien Lotte, Florian Larrue, and Christian Mühl. 2013. Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: Lessons learned from instructional design. *Front. Hum. Neurosci.* 7, September (Jan. 2013), 568.

José Del R. Millán and Josep Mouriño. 2003. Asynchronous BCI and local neural classifiers: An overview of the Adaptive Brain Interface project. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 2 (June 2003), 159–61.

Gernot R. Müller-Putz and Reinhold Scherer. 2008. Better than random? A closer look on BCI results. *Int. J. Bioelectromagn.* 10, 1 (2008), 52–55.

Lennart E. Nacke. 2009. *Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment*. PhD Thesis. Blekinge Institute of Technology. Karlskrona, Sweden.

Lennart E. Nacke and C. A. Lindley. 2008. Flow and immersion in first-person shooters: Measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share (Future Play '08)*. ACM, New York, NY, USA, 81–88. DOI:10.1145/1496984.1496998

Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback game design: Using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, USA, 103–112. DOI:10.1145/1978942.1978958

Carsten Pohl, Wilfried Kunde, Thomas Ganz, Annette Conzelmann, Paul Pauli, and Andrea Kiesel. 2014. Gaming to see: Action video gaming is associated with enhanced processing of masked stimuli. *Front. Psychol.* 5, 70 (2014), 1–9.

Martin Rohm, Alex Kreilinger, Rüdiger Rupp, and Otto Bock Gmbh. 2010. A hybrid-brain computer interface for control of a reaching and grasping neuroprosthesis. *Biomed. Eng. (NY).* 55, 1 (2010), 3–6.

Peter A. M. Ruijten, Cees J. H. Midden, and Jaap Ham. 2011. Unconscious persuasion needs goal-striving: The effect of goal activation on the persuasive power of subliminal feedback. In *Proceedings of the 6th International Conference on Persuasive Technology: Persuasive Technology and Design: Enhancing Sustainability and Health (PERSUASIVE '11)*, Curtis P. Haugtvedt and Agnis Stibe (Eds.). ACM, New York, NY, USA, Article 4, 6 pages. DOI:10.1145/2467803.2467807

N. S. Erlbaum. Schutte, J. M. Malouff, J. C. Post-Gorden, and A. L. Rodasta. 1988. Effects of playing videogames on cheilden's aggressive and other behaviours. *J. Appl. Soc. Psychol.* 18, 5 (1988), 454–460.

Roger W. Schvaneveld and David E. Meyer. 1973. Retreival and Comparison processes in semantic memory. In S. Kornblum (Ed.). *Attention and Performance IV*. New York, NY, USA: Academic Press, Inc., 395–409.

Irina Simanova, Marcel van Gerven, Robert Oostenveld, and Peter Hagoort. 2010. Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS One* 5, 12 (Jan. 2010), e14465.

Tilo Strohbach, Peter A. Frensch, and Torsten Schubert. 2012. Video game practice optimizes executive control skills in dual-task and task-switching situations. *Acta Psychol.* 140 (2012), 13–24.

Yunsick Sung, Kyungeun Cho, and Kyhyun Um. 2012. A development architecture for serious games using BCI (brain computer interface) sensors. *Sensors (Basel).* 12, 11 (January 2012), 15671–88.

Justin Talbot, Bongshin Lee, Ashish Kapoor, and Desney S. Tan. 2009. EnsembleMatrix: Interactive visualization to support machine learning with multiple classifiers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, New York, NY, USA, 1283–1292. DOI:10.1145/1518701.1518895

Desney S. Tan and Anton Nijholt (Eds.). 2010. *Brain-Computer Interfaces - Applying our Minds to Human-Computer Interaction*. Human–Computer Interaction Series. Springer-Verlag London. DOI:10.1007/978-1-84996-272-8.

Eoin Thomas, Matthew Dyson, and Maureen Clerc. 2013. An analysis of performance evaluation for motor-imagery based BCI. *J. Neural Eng.* 10, 3 (Jun. 2013), 031001.

Chi Thanh Vi, Izdihar Jamil, David Coyle, and Sriram Subramanian. 2014. Error related negativity in observing interactive tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14)*. ACM, New York, NY, USA, 3787–3796.

Carmen Vidaurre, Claudia Sannelli, Klaus-Robert Müller, and Benjamin Blankertz. 2010. Machine-Learning Based Co-adaptive Calibration: A Perspective to Fight BCI Illiteracy. In Manuel Graña Romay, Emilio Corchado, and M. Teres. Garcia Sebastian (Eds.). *Hybrid Artificial Intelligence Systems SE - 50*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 413–420.

Marijn van Vliet, Christian Mühl, Boris Reuderink, and Mannes Poel. 2010. Guessing What's on Your Mind: Using the N400 in Brain Computer Interfaces. In Yiyu Yao, Ron Sun, Tomaso Poggio, Jiming Liu, Ning Zhong, and Jimmy Huang (Eds.). *Brain Informatics SE - 17*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 180–191.

Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. 2002. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 6 (June 2002), 767–91.

Received February 2015; revised July 2015; accepted July 2015