

Steady State Visual Evoked Potential Based Computer Gaming – The Maze

Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Arne Robben,
Marijn van Vliet, and Marc M. Van Hulle

KU Leuven, Laboratorium voor Neuro- en Psychofysiologie, Campus Gasthuisberg,
Herestraat 49, B-3000 Leuven, Belgium

{Nikolay.Chumerin,NikolayV.Manyakov,Adrien.Combaz,
Arne.Robben,Marijn.vanVliet,Marc.VanHulle}@med.kuleuven.be

Abstract. We introduce a game, called “The Maze”, as a brain-computer interface (BCI) application in which an avatar is navigated through a maze by analyzing the player’s steady-state visual evoked potential (SSVEP) responses recorded with electroencephalography (EEG). The same computer screen is used for displaying the game environment and for the visual stimulation. The algorithms for EEG data processing and SSVEP detection are discussed in depth. We propose the system parameter values, which provide an acceptable trade-off between the game control accuracy and interactivity.

1 Introduction

With a brain-computer interface (BCI) brain activity is read and used for enabling a subject to interact with the external world, without involving any muscular activity or peripheral nerves. BCI is now widely regarded as one of the most successful applications of the neurosciences and is in a position to significantly improve the quality of life of patients suffering from amyotrophic lateral sclerosis, stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, etc [1].

In this work we consider non-invasive, electroencephalography (EEG)-based BCI method based on the steady-state visual evoked potential (SSVEP). SSVEP is a response recorded from the occipital pole of a brain on the repetitive presentation of visual stimuli (*i.e.*, flickering stimuli). When stimulation is at a sufficiently high rate (starting from 6 Hz), the individual transient EEG responses overlap, leading to a steady state signal: the signal resonates at the stimulus rate and its multipliers [2]. This means that, when a subject is looking at a stimulus flickering at frequency f , one can detect $f, 2f, 3f, \dots$ in the recorded EEG data. Since the amplitude of a typical EEG signal decreases as $1/f$ in the spectral domain [3], the higher harmonics become less prominent. Furthermore, SSVEP is embedded in other on-going brain activity and (recording) noise. Thus, when considering a too small recording interval, erroneous detections are quite likely to occur. To overcome this problem, averaging over several

recording intervals [4], or recording over longer time intervals [5] are often used for increasing the signal-to-noise ratio (SNR) in the spectral domain. Finally, in order to increase the usability and the information transfer rate of the SSVEP-based BCI, the user should be able to select one of several commands, which means that the system should be able to reliably detect several (n_f) frequencies f_1, \dots, f_{n_f} . This makes the frequency detection problem more complex, calling for an efficient signal processing and decoding algorithm.

BCIs were initially aimed for medical purposes, but currently they attract a lot of attention from the entertainment community [6], since they can be used as a new interface, *e.g.*, for mind-controlled games, or for remotely controlling devices. Several studies on SSVEP BCI gaming were published during the last few years [7,8].

In this paper, we present a novel BCI SSVEP game, which achieves good performance thanks to an appropriate detection algorithm combined with spatial filtering. We also discuss some necessary modifications to the game strategy, which can make this brain game more easy to use and more attractive.

2 Methods

2.1 EEG Data Acquisition

The EEG recordings were performed using a prototype of an ultra low-power 8-channels wireless EEG system, which was developed by imec¹, and built around their ultra-low power 8-channel EEG amplifier chip [9]. The data are transmitted at a sampling rate of 1000 Hz, for each channel. We used an electrode cap with large filling holes and sockets for mounting of active Ag/AgCl electrodes (Acti-Cap, Brain Products). The recordings were made with eight electrodes located on the occipital pole (covering the primary visual cortex), namely at positions P3, Pz, P4, PO9, O1, Oz, O2, PO10, according to the international 10–20 electrode placement system. The reference electrode and ground were placed on the left and right mastoids, respectively.

The raw EEG signals are filtered above 3 Hz, with a fourth order zero-phase digital Butterworth filter, so as to remove the DC component and the low frequency drift. A notch filter is also applied to remove the 50 Hz powerline interference.

2.2 Calibration Stage

The game uses only four commands for navigating the avatar through the maze: “left”, “up”, “right” and “down”, hence, four stimulation frequencies are needed. During our preliminary experiments, we noticed that the optimal set of stimulation frequencies is very subject dependent. This motivated us to introduce a calibration stage, preceding the actual game play, for locating the frequency band, consisting of four frequencies, that evoke prominent SSVEP responses in

¹ <http://www.imec.be>

the subject’s EEG signal. To this end, we propose a “scanning” procedure, consisting of several blocks. In each block, the subject is visually stimulated for 15 seconds by a flickering screen ($\approx 28^\circ \times 20^\circ$), after which a black screen is presented for 2 seconds. The number of blocks in the calibration stage is defined by the number of available stimulation frequencies. We have used a laptop with a bright 15,4" LCD screen with a 60 Hz refresh rate. In order to arrive at a visual stimulation with stable frequencies, we show an intense stimulus for k frames, and a less intense stimulus for the next l frames, hence, the flickering period of the stimulus is $k + l$ frames and the corresponding stimulus frequency is $r/(k + l)$, where r is the screen’s refresh rate. Using this simple strategy, one can stimulate the subject with the frequencies that are dividers of the screen refresh rate: 30 Hz (60/2), 20 Hz (60/3), 15 Hz (60/4), and so on. We grouped these frequencies into overlapping bands, for which each band contains four consecutive stimulation frequencies (*e.g.*, band 1: [6 Hz, 6.66 Hz, 7.5 Hz, 8.57 Hz], band 2: [6.66 Hz, 7.5 Hz, 8.57 Hz, 10 Hz], and so on). After stimulation, we visually analyze the spectrograms of the recorded EEG signals, and select the “best” band of frequencies to be used in the game. We have to admit that this frequency selection procedure is subjective, and probably not optimal, calling for an automated procedure.

2.3 Spatial Filtering

Following the minimum energy combination method proposed in [10], we use a spatial filter designed in the following way: a linear combination of the channels is sought that decreases the noise level of the resulting weighted signals at the specific frequencies we want to detect (namely, the frequencies of the oscillations evoked by the periodically flickering stimuli, and their harmonics). This can be done in two steps. In the first step, all information related to the frequencies of interest must be eliminated from the recorded signals. The resulting signals contain only information that is “uninteresting” in the context of our application, and, therefore, could be considered as noise components of the original signals. In the second step, we look for a linear combination that minimizes the variance of the weighted sum of the “noisy” signals obtained in the first step. Eventually, we apply this linear combination to the original signals, resulting in signals with a lower level of noise.

The first step can be done by subtracting from the EEG signal all the components corresponding to the stimulation frequencies and their harmonics. Formally, this can be done in the following way. Let us consider the input signal, sampled over a time window of duration T with sampling frequency F_s , as a matrix \mathbf{X} with channels in columns and samples in rows. Then, one needs to construct a matrix \mathbf{A} , which should have the same number of rows as \mathbf{X} and as the number of columns twice the number of all considered frequencies (including harmonics). For a given time instant t_i (corresponding to the i -th sample in \mathbf{X}) and frequency f_j (from the full list of stimulation frequencies including the harmonics), the corresponding elements $a_{i,2j-1}$ and $a_{i,2j}$ of the matrix \mathbf{A} are computed as $a_{i,2j-1} = \sin(2\pi f_j t_i)$ and $a_{i,2j} = \cos(2\pi f_j t_i)$. For example,

considering only $n_f = 2$ frequencies with their $N_h = 2$ harmonics and a time interval of $T = 2$ seconds, sampled at $F_s = 1000$ Hz, the matrix \mathbf{A} would have $2 \times n_f \times (1 + N_h) = 2 \times 2 \times 3 = 12$ columns and $T \times F_s = 2000$ rows. The most “interesting” components of the signal \mathbf{X} can be obtained from \mathbf{A} by a projection determined by the matrix $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Using $\mathbf{P}_\mathbf{A}$ the original signal without the “interesting” information is estimated as $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{P}_\mathbf{A} \mathbf{X}$. Those remaining signals $\tilde{\mathbf{X}}$ can be considered as noise components of the original signals (*i.e.*, the brain activity not related to the visual stimulation).

In the second step, we use an approach based on Principal Component Analysis (PCA) to find a linear combination of the input data for which the noise variance is minimal. A PCA transforms a number of possibly correlated variables into uncorrelated ones, called principal components, defined as projections of the input data onto the corresponding principal vectors. By convention, the first principal component captures the largest variance, the second principal component the second largest variance, and so on. Given that the input data comes from the previous step, and contains mostly noise, the projection onto the last principal component direction is the desired linear combination of the channels, *i.e.*, one that reduces the noise in the best way (*i.e.*, making the noise variance minimal).

The conventional PCA approach estimates the principal vectors as eigenvectors of the covariance matrix $\Sigma = E\{\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\}$, where $E\{\cdot\}$ denotes the statistical expectancy². Since the considered EEG signal has 8 channels, Σ has size 8×8 , is positive semidefinite and, therefore, it is possible to find a set of 8 orthonormal eigenvectors (represented as columns of a matrix V), such that $\Lambda = V \Sigma V^T$, where Λ is a diagonal matrix of the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_8 \geq 0$. Then, the K last (smallest) eigenvalues are selected such that K is maximal, and $\sum_{k=1}^K \lambda_{9-k} / \sum_{j=1}^8 \lambda_j < 0.1$ is satisfied. The corresponding K eigenvectors, arranged as columns of a matrix V_K , specify a linear transformation that efficiently reduces the noise power in the signal $\tilde{\mathbf{X}}$. The same noise-reducing property of V_K is valid for the original signal \mathbf{X} . Assuming that V_K would reduce the variance of the noise more than the variance of the signal of interest, the signal that is spatially filtered in this way, $\mathbf{S} = V_K \mathbf{X}$, would have greater (or, at least, not smaller) SNR [10].

2.4 Classification

The straight-forward approach to select one frequency (among several possible candidates) present in the analyzed signal is based on a direct analysis of the signal power function $P(f)$ that is defined as follows:

$$P(f) = \left(\sum_t s(t) \sin(2\pi ft) \right)^2 + \left(\sum_t s(t) \cos(2\pi ft) \right)^2,$$

² Since the original signal is high-pass filtered above 3 Hz, the DC component is removed and, therefore, the filtered data are centered (the mean is close to zero).

where $s(t)$ is the signal after spatial filtering. Note that the right-hand part of this equation is the squared Discrete Fourier Transform magnitude at the frequency of interest [10]. The “winner” frequency f^* can then be selected as the frequency with maximal (among all considered frequencies f_1, f_2, \dots, f_{n_f}) power amplitude:

$$f^* = \arg \max_{f_1, \dots, f_{n_f}} P(f).$$

Unfortunately, in our case, this direct method is not applicable due to the nature of the EEG signal: the corresponding power function decreases (similarly to $1/f$) with increasing f . In this case, the true dominant frequency could have an power amplitude less than the other considered lower frequencies. In [5] it was shown that the SNR does not decrease with increasing frequency, but remains nearly constant. Relying on this finding, one can select the “winner” frequency as the one for which the SNR is maximal, $P(f)/\sigma(f)$, where $\sigma(f)$ is an estimation of the noise power for frequency f .

The noise power estimation is not a trivial task. One way to do this is to record extra EEG data from the subject, without visual stimulation. In this case, the power of the considered frequencies in the recorded signal should correspond to the noise level. Despite its apparent simplicity, this method has at least two drawbacks: 1) an extra (calibration) EEG recording session is needed, and 2) the noise level changes over time and the pre-estimated values could significantly deviate from the actual ones. To overcome these drawbacks, we need an efficient on-line method of noise power estimation. As a possible solution, one can try to approximate the desired noise power $\sigma(\tilde{f})$ for a frequency of interest \tilde{f} using values of $P(f)$ from a close neighborhood $O(\tilde{f})$ of the considered frequency \tilde{f} . A simple averaging $\sigma(\tilde{f}) \approx E\{P(f)\}_{f \in O(\tilde{f}) \setminus \tilde{f}}$ produces unstable (jittering) estimates if the size of the neighborhood $O(\tilde{f})$ is small. Additionally, a large neighborhood could contain several frequencies of interest that could bias the estimate of $\sigma(\tilde{f})$.

In our work, we have used an approximation of noise based on an autoregressive modeling of the data, after excluding all information about the flickering, *i.e.*, of signals $\tilde{\mathbf{S}} = V_K \tilde{\mathbf{X}}$ (see previous subsection). The rationale behind this approach is that the autoregressive model can be considered as a filter (working through convolution), in terms of ordinary products between the transformed signals and the filter coefficients in the frequency domain. Since we assume that the prediction error in the autoregressive model is uncorrelated white noise, we have a flat power spectral density for it with a magnitude that is a function of the variance of the noise. Thus, the Fourier transformations of the regression coefficients a_j (estimated, for example, with the use of the Yule-Walker equations) show us the influence of the frequency content of particular signals on the white noise variance ($\tilde{\sigma}$). By assessing such transforms, we can obtain an approximation of the power of the signal $\tilde{\mathbf{S}}$.

More formally, we have:

$$\sigma(f) = \frac{\pi T}{4} \frac{\tilde{\sigma}^2}{|1 - \sum_{j=1}^p a_j \exp(-2\pi i j f / F_s)|},$$

where T is the length of the signal, $i = \sqrt{-1}$, p is the order of the regression model and F_s is the sampling frequency. Since for the detection of each stimulation frequency, we use several channels and several harmonics, we could combine separate values of the SNR as:

$$Q(f) = \frac{1}{K(N_h + 1)} \sum_{k=1}^K \sum_{h=1}^{N_h+1} P_k(hf) / \sigma_k(hf),$$

where K is the dimensionality of the signal \mathbf{S} and $(N_h + 1)$ is the number of the multipliers of the considered frequency f (one fundamental frequency plus its N_h harmonics).

The “winner” frequency f^* is defined as the frequency with the largest index $Q(\cdot)$ among all frequencies of interest:

$$f^* = \arg \max_{f_1, \dots, f_{n_f}} Q(f).$$

2.5 Game Design and Implementation

We have developed a SSVEP-based BCI game “The Maze”, in which the player can control an avatar in a simple maze-like environment. The task is to navigate the avatar (depicted as Homer Simpson’s head) to the target (*i.e.*, a donut) through the maze (see Fig. 1). The game has several pre-defined levels of increasing complexity. A random maze mode is also available. The player can control the avatar by looking at flickering arrows (showing the direction of the avatar’s next move) placed in the periphery of the maze. Each arrow is flickering with its own unique frequency taken from the selected frequency band (see Section 2.2). The selection of the frequencies can be predefined or set according to the player’s preferences.

The game is implemented in Matlab as a client-server application and can run either in parallel Matlab mode (as two labs) or on two Matlab sessions started as separate applications. The server part is responsible for the EEG data acquisition, processing and classification. The client part is responsible for the game logic, user interface and rendering. The client-server communication is implemented using sockets and due to a minimal data transfer rate (during the game only commands are sent from the server to the client) it can work over a regular network, allowing also (optionally) to run the game on two different computers. For the accurate (in terms of timing) visualization of the flickering stimuli, we have used Psychtoolbox 3 (<http://psychtoolbox.org>).

To reach a decision, the server needs to analyze the EEG data acquired over the last T seconds. In the game, T is one of the tuning parameters (must be set

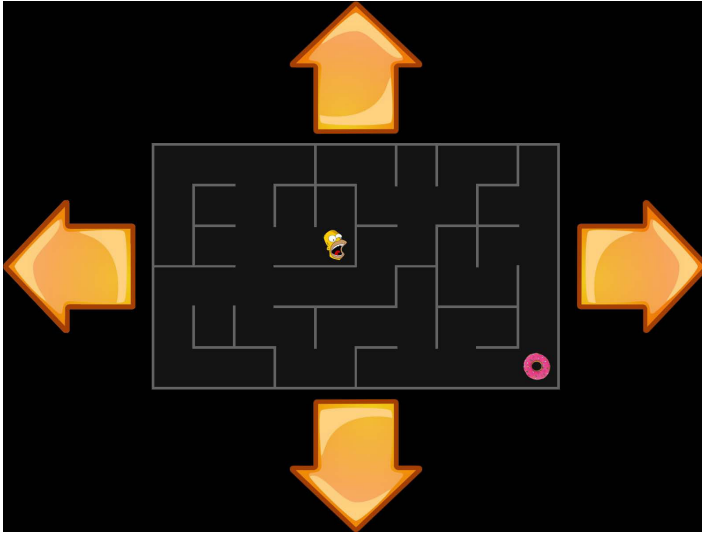


Fig. 1. Snapshot of “The Maze” game

before the game starts), which controls the game latency. Decreasing T makes the game more responsive, but in the same time it makes the interaction less accurate, resulting in wrong navigation decisions. By default, a new portion of the EEG data is collected every 200 ms. The server analyzes the new (updated) data window and detects the dominant frequency using the method described above. The command corresponding to the selected frequency is sent to the client also every 200 ms, thus, the server’s update frequency is 5 Hz. The final decision (the command that is executed) is made by the client using the history of the last m frequency detections: if in the queue of the last m detected frequencies there is a frequency with more than $m/2$ occurrences, then this frequency is considered to be the “final winner”, otherwise no decision is made.

As mentioned above, the game control has an unavoidable time lag. In order to “hide” this latency, we let the avatar change its navigation direction only in so-called decision points: as the avatar starts to move, it will not stop until it reaches the next decision points on its way. This allows the player to use this period of “uncontrolled avatar movement” for planning (by looking on appropriate flickering arrow) the next navigation direction. By the time the avatar reaches the next decision point, the EEG data window, which is to be analyzed, already contains the SSVEP response corresponding to the new navigation direction.

2.6 Influence of Window Size and Decision Queue Length on Accuracy

To assess the best combination of the window size T and the decision queue length m , we have studied their influence on the classification accuracy. Six

healthy subjects (all male, aged 24–34 with average age 28.3, four righthanded, one lefthanded and one bothhanded) participated in the experiment. Only one subject had prior experience with SSVEP-based BCI. For each subject, several sessions with different stimulation frequency sets were recorded, but we present the results only for those sessions, for which the stimulation frequencies coincide with the ones that are determined with the calibration stage. Each subject was presented with a specially designed level of the game, and was asked to consequently look at each one of four flickering arrows for 20 seconds followed by 10 seconds of rest, so the full round of four stimuli (flickering arrows) was $4 \times (20 + 10) = 120$ seconds. The stimulus to attend to was marked with the words “look here”. Each recording session consisted of two rounds and, thus, lasted 4 minutes. The recorded EEG data were then analyzed off-line using exactly the same mechanism as in the game: for each position of the sliding window (of size T) the detected frequency was pushed in the queue (of length m), and the final decision was based on reaching more than 50% of the votes. Due to the design of the experiment, the true winner frequency is known for each moment of time, which enables us to estimate the accuracy.

3 Results and Discussion

The results of the experiment described in Section 2.6 are shown in Table 1. With the accuracy of the frequency classification we mean the ratio of the correct decisions with respect to all decisions made by the classifier. Note, that the chance level of accuracy in this experiment is 25%.

From Table 1 it can be seen that, in general, the longer queues of the decision making mechanism lead to a better accuracy of the game control. The drawback of the longer queues is an additional latency. To reduce the later, the server’s update frequency (the actual one is 5 Hz) can be increased. This, in turn, increases the computational load (mostly on the server part).

Based on our experience (also supported by the data from Table 1), we can recommend to use the window size $T = 3$ and the queue length $m = 5$ (or more) as default values for an acceptable gameplay.

Unfortunately, the information transfer rate (ITR) commonly used as a performance measure for BCIs, is not relevant for the game, at least in its actual form. By design, the locations of the decision points depend on the (randomly generated) maze, and, therefore, the decisions themselves are made at an irregular rate, which, in turn, does not allow for a proper ITR estimation.

A few more issues concerning the visual stimulation and the game design need to be discussed. Even though the visual stimulation in the calibration stage (one full-screen stimulus) differs from the one used in the game (four simultaneously flickering arrows, see Figure 1), we strongly believe that the frequencies selected in such a way are also well suited for the game control. This belief has been indirectly supported during our experiments (see Section 2.6): the frequency sets, different from the ones selected during the calibration stage, in most cases yield less accurate detections.

Table 1. Classification accuracy as a function of window size T and decision queue length m

T (s)	m	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
1	1	57.14%	91.96%	75.22%	53.35%	49.78%	47.32%
	3	58.80%	93.06%	78.24%	52.31%	49.54%	48.38%
	5	59.13%	94.71%	81.49%	53.61%	50.00%	48.08%
2	1	70.83%	99.51%	88.97%	69.36%	64.71%	53.92%
	3	72.70%	100.00%	89.03%	69.39%	65.56%	54.34%
	5	73.14%	100.00%	89.36%	71.01%	66.22%	54.52%
3	1	74.18%	100.00%	92.66%	81.79%	69.84%	62.50%
	3	75.28%	100.00%	92.05%	82.39%	71.31%	62.50%
	5	76.19%	100.00%	92.26%	82.74%	72.02%	62.20%
4	1	73.17%	100.00%	94.82%	86.59%	70.43%	63.11%
	3	74.68%	100.00%	94.55%	87.18%	70.19%	63.14%
	5	75.68%	100.00%	94.93%	89.19%	70.95%	63.51%
5	1	65.28%	100.00%	94.10%	88.89%	65.63%	63.89%
	3	65.81%	100.00%	94.49%	88.97%	65.07%	62.87%
	5	65.63%	100.00%	93.36%	89.45%	64.45%	63.28%

One of the drawbacks of SSVEP-based BCIs with dynamic environment and fixed locations of stimuli is the frequent change of the subject's gaze during the gameplay, which leads to a discontinuous visual stimulation. To avoid this, we introduced an optional mode where the stimuli (arrows) are locked close to the avatar and move with it during the game, which might make the game more comfortable to play.

Several subjects have noticed that the textured stimuli are easier to concentrate on than the uniform ones. Some of our subjects preferred the yellow color of the stimuli to the white color, which partially might be explained by a characteristic feature of the yellow light stimulation: it elicits an SSVEP response of a strength that is less dependent on the stimulation frequency than other colors [11].

BCI-based gaming is research direction that is still in its infancy, and still a lot of issues to be tackled before it could become accepted in the gaming community. All these issues, including the ones discussed above, clearly indicate the necessity of further BCI research, in general, and the development of suitable applications for interactive entertainment, in particular.

Acknowledgments. NC is supported by IST-2007-217077, NVM is supported by research grant GOA 10/019, AC and AR are supported by IWT doctoral grants, MvV is supported by G.05809, MMVH is supported by PFV/10/008, CREA/07/027, G.0588.09, IUAP P6/29, GOA 10/019, IST-2007-217077, and the SWIFT prize of the King Baudouin Foundation of Belgium. The authors grateful to Refet Firat Yazicioglu, Tom Torfs and Cris Van Hoof from imec, Leuven, Belgium, for providing the wireless EEG system.

References

1. Mak, J.N., Wolpaw, J.R.: Clinical Application of Brain-Computer Interface: Current State and Future Prospects. *IEEE Rev. Biomed. Eng.* 2, 187–199 (2009)
2. Herrmann, C.S.: Human EEG Responses to 1–100 Hz Flicker: Resonance Phenomena in Visual Cortex and Their Potential Correlation to Cognitive Phenomena. *Exp. Brain Res.* 137, 346–353 (2001)
3. Allegrini, P., Menicucci, D., Bedini, R., Fronzoni, L., Gemignani, A., Grigolini, P., West, B.J., Paradisi, P.: Spontaneous brain activity as a source of ideal $1/f$ noise. *Phys. Rev. E* 80, 061914 (2009)
4. Manyakov, N.V., Chumerin, N., Combaz, A., Robben, A., Van Hulle, M.M.: Decoding SSVEP Responses Using Time Domain Classification. In: International Conference on Fuzzy Computation and 2nd International Conference on Neural Computation, pp. 376–380 (2010)
5. Wang, Y., Wang, R., Gao, X., Hong, B., Gao, S.: A practical VEP-based brain-computer interface. *IEEE TNSRE* 14(2), 234–239 (2006)
6. Nijholt, A., Plass-Oude Bos, D., Reuderink, B.: Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Comp.* 1(2), 85–94 (2009)
7. Martinez, P., Bakardjian, H., Cichocki, A.: Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm. *Computational Intelligence and Neuroscience* vol. 2007, article ID 94561 (2007)
8. Lalor, E.C., Kelly, S.P., Finucane, C., Burke, R., Smith, R., Reilly, R.B., McDarby, G.: Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment. *EUEASIP J. on Appl. Sign. Proc.* 19, 3156–3164 (2005)
9. Yazicioglu, R.F., Torfs, T., Merken, P., Penders, J., Leonov, V., Puers, R., Gyssels, B., Van Hoof, C.: Ultra-low-power biopotential interfaces and their applications in wearable and implantable systems. *Microel. J.* 40(9), 1313–1321 (2009)
10. Friman, O., Volosyak, I., Gräser, A.: Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE TBE* 54(4), 742–750 (2007)
11. Regan, D.: An effect of stimulus colour on average steady-state potentials evoked in man. *Nature* 210(5040), 1056–1057 (1966)