Subject-Adaptive Steady-State Visual Evoked Potential Detection for Brain-Computer Interface

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

Laboratory for Neurophysiology, K.U.Leuven, Herestraat 49, bus 1021, 3000 Leuven, Belgium
{Nikolay.Chumerin, NikolayV.Manyakov, Adrien.Combaz, Arne.Robben,
Marijn.vanVliet, Marc.VanHulle}@med.kuleuven.be

Abstract – We report on the development of a four command Brain-Computer Interface (BCI) based on steady-state visual evoked potential (SSVEP) responses detected from human electroencephalograms (EEGs). The proposed system combines spatial filtering, feature extraction and selection, and a classifier. Two types of classifiers were compared: one based on equal treatment of all harmonics in all EEG channels and the second based on preliminary training resulting in a weighted treatment of the harmonics. Results from six healthy subjects are evaluated.

Keywords – SSVEP; EEG; BCI; decoding

I. INTRODUCTION

A brain-computer interface (BCI) is a system that enables communication solely based on brain activity without involving any muscular activity. By establishing a communication and/or control channel between a person and an external device (e.g., computer or wheelchair), BCIs can significantly improve the life quality of people with serious motor function problems (i.e., patients suffering from amyotrophic lateral sclerosis, stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, etc [1], [2]).

Brain-computer interfaces are either invasive [3]–[5] or noninvasive [6]–[8]. The invasive ones use recordings made intracortically (local field potentials and action potentials) or from the surface of the brain (electrocorticogram), whereas the noninvasive ones mostly employ electroencephalograms (EEGs) recorded from the subject's scalp.

Several noninvasive methods have been proposed in the literature. The one we consider in this paper, is based on the *steady-state visual evoked potential* (SSVEP). This type of BCI relies on the psychophysiological properties of EEG brain responses recorded from the occipital pole during the periodic presentation of identical visual stimuli (i.e., flickering stimuli). When the periodic presentation is

at a sufficiently high rate (≥ 6 Hz), the individual transient visual responses overlap, leading to a *steady state* signal: the signal resonates at the stimulus rate and its multipliers [9]. This means that, when the subject is looking at stimuli flickering at the frequency f_1 , the frequencies f_1 , $2f_1$, $3f_1$, $\dots$ can be detected in the Fourier transform of the EEG signal recorded from the occipital pole. Since the amplitude of a typical EEG signal is proportional to $1/f$ in the spectral domain, the higher harmonics become less prominent. Furthermore, the fundamental harmonic f_1 is embedded into other on-going brain activity and (recording) noise. Thus, when considering a small recording interval, it is quite possible to erroneously detect a (wrong) frequency f_1 . To overcome this problem, averaging over several time intervals [10], recording over longer time intervals [11], and/or preliminary training [4], [12], [13] are often used for increasing the signal-to-noise and detectability of responses. Finally, in order to increase the usability and 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* SSVEP-induced frequencies $f_1, \dots, f_n$ (corresponding to these commands) in the EEG data. This makes the SSVEP detection problem more complex, and requires an efficient signal processing and decoding algorithm.

II. METHODS

A. EEG Data Acquisition

The EEG recordings for our experiments were performed using a prototype of an ultra low-power 8-channels *wireless* EEG system, which consists of two parts: an amplifier coupled with a *wireless* transmitter and a USB stick receiver. This system was developed by imec¹, and built around their ultra-low power 8-channel EEG amplifier chip [14]. The data is sampled and transmitted at sample rate of $F_s = 1000$ Hz for each channel. We used an electrode cap with large filling holes, and sockets for mounting active Ag/AgCl electrodes (ActiCap, Brain Products). The recordings were made with eight

NC is supported by IST-2007-217077, NVN is supported by the 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 King Baudouin Foundation of Belgium (SWIFT prize).

¹ Interuniversity Microelectronics Centre (imec), <http://www.imec.be>

Figure 1. Scheme of SSVEP scanning procedure during the calibration stage.

electrodes located on the occipital pole (covering 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 and the ground electrodes 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.

B. Calibration Stage

During the preliminary experiments, we noticed that the optimal stimulation frequency is very subject dependent. This motivated us to introduce a *calibration* stage. The goal of this stage is to detect the stimulation frequencies that can be further robustly detected in the subject's EEG data. Since we use BCI system with four commands, we have to select four best detectable frequencies during the calibration stage. To this end, we propose a "scanning" procedure, which consists of several blocks. In each block the subject is visually stimulated for 15 seconds by a flickering square shown in the center of the screen, after which a black screen is presented for 2 second rest (see Fig. 1). The number of blocks in the calibration stage is defined by the number of stimulation frequencies. For the experiments we have used a laptop with a bright 15,4" LCD screen working at 60 Hz refresh rate. In order to perform a visual stimulation with stable frequencies, we show an intense stimulus for k frames and a not intense stimulus for the next l frames, so 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 ($r = 60$ Hz). Using this simple strategy, one can stimulate the subject with the frequencies which are dividers of the screen refresh rate: 30 Hz (60/2), 20 Hz (60/3), 15 Hz (60/4) and so on. After stimulation, we visually analyze the spectrograms of the recorded EEG signals (see, for example, Fig. 2), and select the four most salient frequencies.

III. BCI SYSTEM

Our SSVEP BCI system could be represented as a Waterfall-like diagram (see Fig. 3). It consists of three blocks: spatial filtering (which finds a linear combination of the EEG channels in such way, that amplitudes in the frequencies of interest become more salient [see

Figure 2. Spectrogram of EEG recordings from electrode Oz for subject 3, based on a visual stimulation at frequencies 60/4, ..., 60/9 Hz. Note that not only the fundamental frequencies, but also their harmonics are visible. For this subject, frequencies 10 (60/6), 8.57 (60/7), 7.5 (60/8) and 6.67 (60/9) Hz were selected for the BCI control stimuli.

Sec. III-A)], feature extraction and selection (which estimates the signal-to-noise ratio (SNR) coefficients for the selected frequencies in the frequency domain [see Sec. III-B]) and classifier (which detects the frequency of the stimulus the subject looking at [see Sec. III-C]).

A. Spatial Filtering

Similarly to [15], we use a spatial filter to find a linear combination of the channels, which would decrease the noise level of the resulting weighted signals at the specific frequencies we want to detect (oscillations evoked by the flickering stimuli and their harmonics). This can be done in two steps. First, all information related to the frequencies of interest must be eliminated from the recorded signals. The resulting signals contain only information that is "not interesting" in the context of our application, and, therefore, could be considered as the noise components of the original signals. At the second step, we look for a linear combination which would minimize the variance of the weighted sum of the "noisy" signals obtained in the first step. We finally 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 input data 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

Figure 3. Waterfall-like scheme of BCI decoding system.

sampling frequency F_s as a matrix $\mathbf{X}$ with channels in columns and samples in rows. Then, one needs to construct a matrix A , which should have the same number of rows as $\mathbf{X}$ and number of columns twice as more than the number of all considered frequencies (including harmonics). For a time instant t_i (corresponding to the i -th sample in $\mathbf{X}$) and a frequency f_j (from the full list of frequencies including the harmonics) the corresponding elements $a_{i,2j-1}$ and $a_{i,2j}$ of the matrix 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 on a time interval of duration $T = 2$ seconds, sampled with $F_s = 1000$ Hz, the matrix 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 A by a projection specified by the matrix $P_A = A(A^T A)^{-1} A^T$. With P_A it is easy to estimate the original signal *without* the "interesting" information as $\tilde{\mathbf{X}} = \mathbf{X} - P_A \mathbf{X}$. Those remaining signals $\tilde{\mathbf{X}}$ can be considered as the noise components of the original signals (*i.e.*, the brain activity not related to the visual stimulation) and now we need to linearly combine these signals to minimize the variance of the resulting (noisy) signal.

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 corresponding *principal vectors*. By convention, the first principal component captures the largest variance, the second principal component the second largest variance, and so on. Consequently, the last principal vector specifies the direction of the smallest variance of the input data. Considering that the input data come from the previous step and contain mostly noise, the projection onto the last principal component direction (direction of the eigenvector for the signals' covariance matrix $\tilde{\mathbf{X}}$ with smallest eigenvalue) is the desired linear combination of the channels, *i.e.*, that reduces the noise in the best way (making the noise variance minimal). Hence, for the eigenvalues ranked in descending order, we select only the K last (smallest) eigenvalues such that K is maximal and $\sum_{i=1}^K \lambda_{9-i} / \sum_{j=1}^8 \lambda_j < 0.1$ is satisfied. The corresponding k eigenvectors, arranged as columns of a matrix V_K , specify a linear transformation that most 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 spatially filtered in this way signal $\mathbf{S} = V_K \mathbf{X}$ would have greater (or, at least, not smaller) SNR. This assumption has been proven experimentally.

B. Feature extraction and selection

As a features for the classification, we cannot use *power spectral density* (PSD) amplitudes $P(f)$ by themselves, since PSD is inversely proportional to the frequency f . In this case, the true dominant frequency could have an PSD amplitude less than the other considered frequencies. But in [11] it was shown that the signal-to-noise ratio (SNR) does not decrease with increasing frequency, but remains nearly constant. Relying on this assumption, one can select the "winner" frequency as a frequency with the maximal SNR, which is defined as $P(f)/\sigma(f)$, where $\sigma(f)$ is an estimate of the noise power for frequency f . The power of the frequency f , can be estimated as

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

where $s(t)$ is the signal after the spatial filtering.

In this work, following [15], 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 reasoning 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 onto the white noise variance. Thus, by assessing such transforms, we can obtain an approximation of the PSD of our signal. 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, $\tilde{\sigma}$ is an estimate of the variance of white noise, $i = \sqrt{-1}$, p is the order of the regression model and F_s is the sampling frequency.

C. Classification

Since for the detection of each stimulation frequency, we use several channels and several harmonics, we could combine separate values of the SNR as $T(f) = \sum_{i=1}^N \sum_{k=1}^K w_{ik} P_i(kf) / \sigma_i(kf)$, where i is the channel index and k is the harmonic index. The "winner" frequency f^* is defined as the frequency having largest index T among all frequencies of interest $f^* = \arg \max_{f \in \{f_1, \dots, f_n\}} T(f)$.

Normally, equal weight values ($w_{ik} = \frac{1}{NK}$) are used for estimation of $T(f)$ [10], [15]. Thus, SNR at all harmonics are treated equally. But this choice could not be

Figure 4. Snapshot of “The Maze” game.

always convenient. We propose to consider these weights as parameters, by adjusting which the system could be adapted for a particular subject and/or particular recording session of the subject. To train the weights one can re-use data from the calibration stage, where the desired outputs of the classifier are known *a-priori* due to the calibration stage design.

The above mentioned weighting procedure can be represented by a artificial linear neural network. As the input, we used SNR coefficients $P_i(kf)/\sigma_i(kf)$ for every channel and every harmonic. Thus, for eight electrode EEG system and considering the fundamental stimulation frequency and its two harmonics, we have $8 \times 3 = 24$ elements in input vector. As the output $\tilde{T}$, we assign a fixed positive value (+1) for the case, when the input SNRs corresponds to a stimulation frequency, and zero otherwise. Training can be performed using least-square algorithm with additional restrictions on values of the weights (nonnegativity).

Trained in this way network can estimate values $\tilde{T}(f_i)$ for each stimulation frequency f_i , given considered EEG data. The “winner” frequency, again, is selected as the frequency having largest index $\tilde{T}$ among all frequencies of interest f_i , (in the proposed system $i = 1, \dots, 4$).

IV. RESULTS AND DISCUSSION

Our BCI system was implemented as “The Maze” game [16], where a subject can control an avatar in a simple maze-like environment by looking at flickering arrows (showing the direction of avatar’s next move, see Fig. 4) located on the periphery of the maze. Each arrow is flickering with its own unique frequency selected from the set of possible stimulation frequencies (see Section II-B). The selection of the frequencies can be predefined or set according to the player’s preferences. The system 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 (see Fig. 5). 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 visual stimulation. The client-server communication is implemented using sockets and due to minimal data

Figure 5. Client-server architecture of the “The Maze” game.

transfer rate (during the game only commands are sent from the server to the client) can work over a regular network, allowing for running the system on two different computers. For the accurate (in terms of timing) visualization of the flickering stimuli we have used Psychtoolbox².

To reach a decision, the server needs to analyze the EEG data acquired during the last t seconds (correspondingly $T = (F_s \times t)$ samples). In the proposed BCI system, t is one of the tuning parameters (can be set before the game starts) which control the game latency. A new portion of the EEG data arrives every 200 ms. The server analyzes the new (updated) data window and detects the dominant frequency using the method described above and send the correspondent command to the client. To assess the best window size t and decoding method (simple averaging or preliminary training), we have studied the dependency of the classification accuracy from t and the method. Six healthy subjects participated in the experiment. The stimulation frequencies for each subject have been chosen in advance during the calibration stage. Each subject was presented with a specially designed level of “The Maze” 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 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 four minutes. The EEG data recorded during the second round were then analyzed off-line using exactly the same mechanism as in the BCI system. In the case of training mode, first round was used for training. By design, the true winner frequency is known for each moment of time, which enables us to estimate the accuracy. The results of this experiment are shown in Table I.

Some issues concerning the visual stimulation 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 4), we strongly believe that the frequencies selected in such a way are also well suited for the game control. This belief has been indirectly

² <http://psychtoolbox.org>

TABLE I. CLASSIFICATION ACCURACY AS A FUNCTION OF WINDOW SIZE t AND METHOD OF SNR WEIGHTING
(A - AVERAGING, T - BASED ON PROPOSED TRAINING).

t (s)	method	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6	Average $\langle \cdot \rangle$	$\langle T \rangle - \langle A \rangle$
1	A	54.17%	41.15%	35.42%	78.65%	69.27%	55.73%	55.73%	3.99%
	T	54.69%	46.88%	43.23%	81.77%	70.83%	60.94%	59.72%	
2	A	59.78%	50.54%	51.09%	93.48%	82.07%	66.30%	67.21%	9.78%
	T	79.35%	58.70%	63.04%	92.93%	86.96%	80.98%	76.99%	
3	A	69.19%	62.79%	54.07%	94.77%	88.95%	69.19%	73.16%	9.30%
	T	84.30%	68.60%	61.63%	99.42%	94.19%	86.63%	82.46%	
4	A	77.44%	67.07%	52.44%	95.12%	90.24%	75.61%	76.32%	6.30%
	T	86.59%	73.17%	51.83%	100.00%	95.73%	88.41%	82.62%	
5	A	82.89%	69.74%	51.97%	99.34%	96.71%	71.71%	78.72%	5.60%
	T	90.13%	75.66%	57.24%	100.00%	97.37%	85.53%	84.32%	

supported during our experiments: the frequency sets, different from the ones selected during the calibration stage, in most cases yield less accurate detections.

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, one can also introduced an optional mode where the stimuli (arrows) are locked close to the avatar and move with it during the game.

As the features for our decoder, only power spectral densities and estimated SNR's from each channel separately were used. We believe that inclusion of features showing inter-channel relation, such as synchronization [17] or characteristics of propagating waves [18], can also be helpful in improving the decoding performance.

V. CONCLUSION

We have developed four command SSVEP BCI system, which employs time and spatial filtering to increase SNR of the frequencies of interest, feature extraction and selection strategies and two types of SSVEP decoders. From Table I one can see that the proposed (weighted) version of the decoder outperforms the standard (averaged) one by approximately 7% in terms of accuracy.

ACKNOWLEDGMENT

The authors wish to thank Refet Firat Yazicioglu, Tom Torfs, and Chris Van Hoof, from imec in Leuven, for providing us with the wireless EEG system and for their support.

REFERENCES

- [1] J. Mak and J. Wolpaw, "Clinical applications of brain-computer interfaces: current state and future prospects," *Biomedical Engineering, IEEE Reviews in*, vol. 2, pp. 187–199, 2009.
- [2] N. Manyakov, N. Chumerin, A. Combaz, and M. Van Hulle, "Comparison of linear classification methods for P300 Brain-Computer Interface on disabled subjects," in *International Conference on Bio-Inspired Systems and Signal Processing (BIOSIGNALS), Rome, Italy*, 2011, pp. 328–334.
- [3] M. Velliste, S. Perel, M. Spalding, A. Whitford, and A. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature*, vol. 453, no. 7198, pp. 1098–1101, 2008.
- [4] N. Manyakov and M. Van Hulle, "Decoding grating orientation from microelectrode array recordings in monkey cortical area V4," *International Journal of Neural Systems*, vol. 20, no. 2, pp. 95–108, 2010.
- [5] E. Leuthardt, G. Schalk, J. Wolpaw, J. Ojemann, and D. Moran, "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 1, p. 63, 2004.
- [6] N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The thought translation device (TTD) for completely paralyzed patients," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 190–193, 2000.
- [7] N. Chumerin, N. Manyakov, A. Combaz, J. Suykens, R. Yazicioglu, T. Torfs, P. Merken, H. Neves, C. Van Hoof, and M. Van Hulle, "P300 detection based on feature extraction in on-line Brain-Computer Interface," in *Lecture Notes in Computer Science: Vol. 5803/2009. 32nd Annual Conference on Artificial Intelligence. Paderborn, Germany*. Springer, 2009, pp. 339–346.
- [8] B. Blankertz, G. Dornhege, M. Krauledat, K. Müller, and G. Curio, "The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [9] S. Luck, *An introduction to the event-related potential technique*. The MIT Press, Cambridge, Massachusetts, 2005.
- [10] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 10, pp. 1181–1186, 2002.
- [11] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, "A practical VEP-based brain-computer interface," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 234–240, 2006.
- [12] R. de Peralta Menendez, J. Dias, J. Soares, H. Prado, and S. Andino, "Multiclass brain computer interface based on visual attention," in *ESANN2009 proceedings, European Symposium on Artificial Neural Networks, Bruges, Belgium*, 2009, pp. 437–442.
- [13] A. Luo and T. Sullivan, "A user-friendly SSVEP-based brain-computer interface using a time-domain classifier," *Journal of Neural Engineering*, vol. 7, p. 026010, 2010.
- [14] R. Yazicioglu, T. Torfs, P. Merken, J. Penders, V. Leonov, R. Puers, B. Gyselinckx, and C. Van Hoof, "Ultra-low-power biopotential interfaces and their applications in wearable and implantable systems," *Microelectronics Journal*, vol. 40, no. 9, pp. 1313–1321, 2009.
- [15] O. Friman, I. Volosyay, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 4, pp. 742–750, 2007.
- [16] N. Chumerin, N. Manyakov, A. Combaz, A. Robben, M. van Vliet, and M. Van Hulle, "Steady state visual evoked potential based computer gaming - The Maze," in *The 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2011). Genoa, Italy*, 2011.
- [17] N. Manyakov and M. Van Hulle, "Synchronization in monkey visual cortex analyzed with an information-theoretic measure," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, p. 037130, 2008.
- [18] N. Manyakov, R. Vogels, and M. Van Hulle, "Decoding stimulus-reward pairing from local field potentials recorded from monkey visual cortex," *IEEE Transactions on Neural Networks*, vol. 21, no. 12, pp. 1892–1902, 2010.