Steady State Visual Evoked Potential (SSVEP) - based Brain Spelling System with Synchronous and Asynchronous Typing Modes

H. Segers[1], A. Combaz[2], N.V. Manyakov[2], N. Chumerin[2], K. Vanderperren[1], S. Van Huffel[1] and M.M. Van Hulle[2]

[1] ESAT – SCD/SISTA, K.U.Leuven, Kasteelpark Arenberg 10, POBox 2446, 3001 Heverlee, Belgium
[2] Laboratorium voor Neuro- and Psychofysiology, K.U.Leuven, Herestraat 49, POBox 1021, 3000 Leuven, Belgium

Abstract — **The paper presents an EEG-based wireless brain-computer interface (BCI) with which subjects can mind-spell text on a computer screen. The application is based on the detection of steady-state visual evoked potentials (SSVEP) in EEG signals recorded on the scalp of the subject. The performance of the BCI is compared for two different classification paradigms, called synchronous and asynchronous modes.**

Keywords — **brain-computer interface, mind speller, steady-state visual evoked potentials, synchronous and asynchronous spelling**

I. Introduction

Research on *brain-computer interfaces* (BCIs) has witnessed a tremendous development in recent years, and is now widely considered as one of the most successful applications of neuroscience. BCIs can significantly improve the quality of life of patients suffering from amyotrophic lateral sclerosis, stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, etc.

Among the different BCIs, mostly the noninvasive ones (see Fig. 1 for a general overview) received a lot of attention lately, since they mostly employ *electroencephalograms* (EEGs) recorded from the subject's scalp without requiring any surgery. In this paper, we study one such type of BCI, based on the detection of *steady-state visual evoked potential* (SSVEP) responses. This type of BCI relies on the psychophysiological properties of EEG brain responses recorded from the occipital pole during the periodic presentation of a visual stimulus (*i.e.*, flickering stimulus). When the periodic presentation is at a sufficiently high rate (>6 Hz), the individual transient responses overlap, leading to a *steady state* signal: the signal resonates at the stimulus rate and its multipliers [1]. This means that, when the subject is looking at stimuli flickering at frequency f_1, the frequencies f_1, $2f_1$, $3f_1$ … can be detected in the Fourier transform of the EEG signal recorded form the occipital pole. Since the amplitude of a typical EEG signal decreases as *1/f* in the spectral domain, the higher harmonics become less prominent. Furthermore, the fundamental harmonic f_1 is embedded into other, ongoing brain activity and (recording) noise. Thus, when considering a small recording interval, it is quite possible to erroneously detect frequency f_1. To overcome this problem, averaging over several time intervals [2], recording over longer time intervals [3], and/or system training [4–5] are often used to increase the signal-to-noise ratio and the detectability of the targted 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* frequencies $f_1,\ldots, f_n$ (one for each command). This makes the frequency detection problem more complex, and requires efficient signal processing and decoding algorithms.

An SSVEP-based BCI could be build as a system operating in a synchronous or asynchronous mode. The first one assumes that the subject observes a stimulus for a fixed, predefined amount of time, after which the classification is performed. This mode requires either a fixing of the stimulation duration for all subjects' or to perform a preliminary training/calibration to adjust the stimulus duration to each subject individnally. The asynchronous mode assumes that the stimulation and decoding are done in parallel, thus, enabling for a proper classification, given a sufficient amount of data is available.

II. Methods

A. EEG data acquisition

The EEG recordings were performed using a prototype of an ultra low-power 8-channel wireless EEG system. This system was developed by *imec*[1] and built around their ultra-low power eight-channel EEG amplifier chip [6]. Recordings were made with eight electrodes located on the parietal and occipital poles, namely in positions P3, Pz, P4, PO9, O1, Oz, O2, PO10 according to the international 10–20 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

[1] Interuniversity Microelectronics Centre, http://www.imec.be

notch filter is also applied to remove the 50 Hz powerline interference.

B. Experiment design

Eight healthy subjects (aged 24–60 with average 35, two female and six male) participated in the two experiments. In the first experiment the subject had to observe a flickered square (white square on the black background) in the center of the screen. For this, eight different frequencies were used, one following the other. Those frequencies were selected as a dividers of the screen refreshing rate (60 Hz) and were 30, 20, 15, 12, 10, 60/7=8.57, 7.5 and 60/9=6.67 Hz. This set of frequencies has been chosen because of the stimulation method: an *intense* (white square) followed by a *non-intense* (black square) stimulus presented for an integer amount of frames (*e.g.*, 1 frame for the intense- and 1 frame for the non-intense square for a stimulation frequency of 30 Hz). The subject's task is to simply concentrate on the flickering stimuli. The data from this experiment is used when we wish to select the best frequencies, for which the SSVEP can be detected, for a given subject.

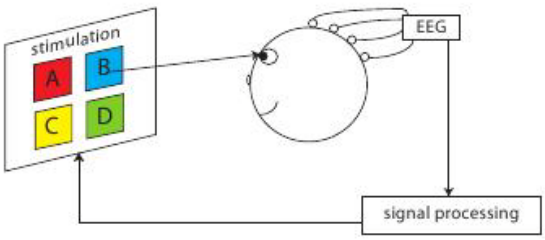

Fig. 1 General overview of the Brain Spelling System.

The second experiment is the actual typing with the spelling device. The subject is presented with a screen with a set of characters arranged in a 8 by 8 matrix. This matrix is divided into four quadrants (sub-matrices of 4 by 4 characters) against a differently colored background. The background of each quadrant is flickering with a particular and unique frequency (selected from the above mentioned set), whereby the subject is allowed to select one group of characters via her/his SSVEP responses while gazing at the correspondent quadrant. After the desired quadrant is selected, it is enlarged in to cover the entire screen and it replaces the initial 8 by 8 matrix. Then, the selection is done and the procedure is repeated: the 4 by 4 matrix is also split into 4 quadrants from which the subject can select only one. Eventually, after three selections, the system detects the character the subject was focusing on. Figure 1 presents the last level in this selection hierarchy.

C. Synchronous and asynchronous modes

For the spelling device two different decoding modes are available: *synchronous* and *asynchronous*. In the synchronous mode the stimulation, signal processing and decoding are sequential: the stimulation lasts for a fixed time Δt, after which the acquired EEG-signals are processed to detect one out of four stimulation frequencies. In the asynchronous mode, all the system components work in parallel: the signal recording, processing and decoding are done during the stimulation phase. For the asynchronous mode, the decoding could be performed on a separate computer, connected to the stimulation computer by mean of a TCP/IP protocol, or on the same computer as one of the concurrent processes. Decoding starts after a short initial pause Δt_p following the beginning of the visual stimulation. During this time the system keeps collecting EEG data. If after Δt_p seconds the collected data allows the classifier to make a "firm" decision, this decision is considered as "final" for this selection stage, and the system goes to the next selection stage. Otherwise, the classifier tries to detect the winner frequency using more data, which have been acquired during a bit longer period $\Delta t_p+\Delta t_c$, where Δt_c is the time needed for the classifier to perform the first classification attempt. The process repeats until the decision is reached or the stimulation time exceeds a threshold Δt_{max} (5 seconds in our case). In the latter case, a most probable classification result is taken (see further).

D. Spatial filtering

For enhancing the decoding accuracy, a feature extraction procedure is applied, which in our case consists of spatial filtering. This means that we search for the optimally weighted linear combination of recording channels, which result in a smaller (or equal) number of signals $S = Y*W$, with an improved signal-to-noise ratio for the frequencies of interest, with W the weight matrix and Y the original EEG signals To estimate W, we opted for the *minimal noise energy* [7] approach, the method of *maximizing the contrast between SSVEP and the noise energy* [7], and for the methods based on the extraction of the *independent components*. The first two methods utilize irrelevant information Y_{ir} which is defined as the result of the subtraction, from the original recordings Y of all information contained in the projections onto the subspace spanned by the sine and cosine transform of all stimulus frequencies and their harmonics. This irrelevant information, considered as noise, has to be minimized to increase the signal-to-noise ratio. Thus, we could look for the weighted combination $Y_{ir}*W$, which minimizes the variances of the resulting signals. The *minimal noise energy* approach uses the principal component analysis to select the directions that correspond to the smallest

eigenvectors of the covariance matrix of Y_{ir}. But the application of the previously determined weights W to the EEG signals can lead also to a decrease in the amplitude of the frequencies of interest. Thus, alternatively, we can look for those weights that increase the signal-to-noise ratio at the frequencies of interest, by maximizing the Rayleigh quotient $\max_w(||Y*w||^2/||Y_{ir}*w||^2)$. This can be done by computing the generalized eigenvalue-decomposition of the matrices Y^T*Y and $(Y_{ir})^T*Y_{ir}$, leading to a *maximum contrast* solution, by taking the directions corresponding to the highest eigenvectors. The idea behind using an independent component analysis (ICA) is based on the assumption that the recorded EEG signals are linear mixtures of independent sources caused by the visual stimulation, the ongoing brain activity and the recording noise. Hence, the application of ICA can lead to the extraction of relevant information. But here a problem arises: which independent component(s) relate to the stimulus activity?

Table 1 Mean classification accuracy for different spatial filtering methods (minimum energy approach, maximum contrast methods and ICA with different numbers of independent components kept)

Method	min energy	max contrast	ICA-8	ICA-7	ICA-6
accuracy	**64%**	63%	63%	61%	59%

E. Classification

Classification was done with the use of $T(f)$, which is an average of the signal-to-noise ratio in the power spectral density (psd) function for all signals remaining after spatial filtering, and all considered harmonics of frequency f (we used 2 harmonics). This statistic was assessed for all possible stimulation frequencies, leading to the selection of the highest value as a "winner" frequency in the synchronous mode. In the asynchronous mode, the "precise" or "firm" classification is done only if the ratio of the highest $T(f)$ to the second highest is greater than some quality threshold Q. Otherwise, the system will require more data for the classification (see Sec. II.C).

The signal-to-noise ratio for a frequency f is estimated as the ratio of the *psd* amplitude at this frequency, during stimulation, and the *psd*, at the same frequency, but with no stimulation. The latter is computed as an approximation of the *psd* when applying the autoregressive method [8] on the signal obtained after subtracting of all relevant information, as was described in Sec. II.D. But this time, the subtraction is applied after spatial filtering.

III. Results

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 (possibly on different systems).

To assess the potential accuracy of the different spatial filtering techniques, we used the data recorded in the first experiment. We tried to decode the frequency the subject was looking each time 5 seconds at one of the all eight stimulation frequencies used in the experiment. This leads to a chance level of 12.5%. The results in Table 1 show that all three spatial filtering strategies lead to almost similar results, while the *minimum noise energy* method performs slightly better, managing to correctly classify 64% of all collected SSVEP-signals. For the decoding performance, without spatial filtering, only 39% of the signals were correctly classified. Including the spatial filtering thus leads to an about 25% increase in detection performance.

We have also determined the best detectable frequency from those eight among all subjects. Based on the data from the first experiment, we have found a peak at 12 Hz (89%). But when choosing an optimal combination of four frequencies (since our online speller has 4 flickering stimuli), out of eight candidates, it is important to not only look for the best individual frequencies, but also to eliminate those that cause the most *false positives*. 12 Hz seems to yield good results, but its usability should be checked for each subject, since it falls into the alpha-range (8–13 Hz). Since this alpha rhythm typically occurs in the EEG, when the subject closes his/her eyes, we could easily obtain misclassifications due to the alpha rhythm.

The second experiment, online typing, was been done with the *minimum noise energy* method, which is the best spatial filtering method according to the study described above. The experiment was performed in the synchronous mode with 5 seconds per selection stage and with the best decoding frequencies selected based on the first experiment. Averaged among all subjects, the typing accuracy was 81%, with the chance level being 100/64=1.5625%. This result shows the potential of our application for a typing device.

To make a qualitative comparison between the synchronous and asynchronous modes, the data recorded with the previous on-line typing method was also subjected to a classification based on an asynchronous decoding. We would like to mention here that this mode also works on-line, and that it was applied in a way that mimics online decoding. Table 2 shows the averaged detection percentages for different initial pauses Δt_p and quality thresholds Q. Additionally, Table 3 shows the corresponding averaged detection times. Note that, in some cells, we have a stimulation and detection time larger than $\Delta t_{max} = 5$ sec, since the table shows the time required for a stimulation *with classification*. The results show that the higher Q, the better the classification results become, but the slower the detection time is. This is as expected because the frequency to be

classified needs to be more pronounsed. This takes more time to achieve, but once this threshold is reached, it is more plausible that the classified SSVEP-frequency is the correct one. Longer initial pauses also lead to better classification results and slower detection times. A possible explanation is that the SSVEP-response is not prominent enough if the initial pause is too short, due to the latency of the responses, or the time required setting a *steady state* mode.

Table 2 Accuracy for different initial pauses Δt_p and quality threshold Q

		Quality threshold Q				
% detected		1.1	1.3	1.5	1.7	1.9
Δt_p [s]	0.5	15%	20%	36%	47%	57%
	1	37%	47%	58%	60%	65%
	1.5	44%	56%	62%	64%	66%

Table 3 Averaged detection time for different initial pause and threshold

		Quality threshold Q				
Avg time [s]		1.1	1.3	1.5	1.7	1.9
Δt_p [s]	0.5	0.55	0.97	2.34	3.41	4.35
	1	1.12	2.25	3.56	4.41	5.12
	1.5	1.74	3.11	4.38	5.20	5.71

Table 4 contains the typing accuracy per subject in the asynchronous mode (Q=1.5 and Δt_p =1.5 s). The first row gives the detection percentages. All subjects manage to achieve near perfect classification results. The second row gives the average detection times. A rather large inter-subject variability can be found here. The third row gives the detection percentages for the eight frequencies, when taking the data of the first experiment. This can be considered as a measure of the *quality* of the SSVEP-response for that person. There is a correlation of 92% between the average detection time and this measure. This reflects a great advantage for the asynchronous classification: all subjects (in this study) can reach almost perfect detection rates, but the classification times for persons with a strong SSVEP-response are a lot shorter. The stimulation time adapts to the specific needs of the subject.

Table 4 Classification results and time per person for 4 command asynchronous typing together with general detection accuracy

	person							
	A	B	C	D	E	F	G	H
% correct	94	100	100	100	94	100	95	100
avg time [s]	2.04	2.66	2.05	2.65	6.36	2.55	5.12	4.86
% general det	74	78	77	74	58	81	63	64

We also made a comparison between synchronous and asynchronous modes based on the theoretical *information transfer rate* (ITR), which tell us how many bits per minute the system can theoretically communicate. It implies that we assume a zero time for changing from one selected target to the next. The ITR averaged over our subjects was used for the assessment, since we wanted to compare the asynchronous with the synchronous one, when the duration of the stimulation was fixed before the experiment, and does not depend on the subject. Hence, we do not consider specially selected stimulation timings for each particular subject, but rather consider the system without any preliminary calibration/training. We can conclude from Table 5, that, in general, the asynchronous mode (Q=1.5 and Δt_p =1.5 s) yields higher ITR's than the synchronous one. Examining the performance of each subject for asynchronous typing, we see that the theoretical ITR's, which can be achieved with a 4 target system, are between 17,57 and 59,16 bits/min. If we use 8 target stimuli, in the asynchronous mode, the ITR drops to 29.4 bits/min.

Table 5 Averaged ITR [bits/min] for different modes and 4 targets

Synchronous with different stimulation durations					Asynch
1s	2s	3s	4s	5s	
35.7	33.4	28.8	22.9	19.0	38.2

IV. CONCLUSIONS

Due to the large inter-subject variability, no pre-selected combination of parameters could be found that leads to an optimal detection percentage or ITR. As such, either a calibration/training stage to tune the system to a particular subject, or an asynchronous classification is needed. In this way, an average ITR of 38 bits/min can be achieved.

REFERENCES

1. Luck S.J (2005) An introduction to the event-related potential technique. MIT Press Cambridge.
2. Cheng M., Gao X., Gao S., and Xu D. (2002) Design and implementation of a brain-computer interface with high transfer rates, IEEE TBE, 49(10): 1181-1186.
3. Wang Y., Wang R., Gao X., Hong B., and Gao S. (2006) A practical VEP- based brain-computer interface, IEEE TNSRE, 14(2): 234-240.
4. Manyakov N.V., Chumerin N., Combaz A., Robben A., and Van Hulle M.M. (2010) Decoding SSVEP responses using time domain classification, Proceedings of the International Conference on Fuzzy Computation and 2nd International Conference on Neural Computation, pp. 376-380.
5. Sergio P., Luca M., Turconi A.C., and Andreoni G. (2009) A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication, Computational Intelligence and Neuroscience 864564.
6. Yazicioglu R.F., Merken P., Puers R., Van Hoof C. (2006) Low-power lownoise 8-channel EEG front-end ASIC for ambulatory acquisition systems, Proceedings of the 32nd European Solid-State Circuits Conference, pp. 247–250.
7. Friman O., Volosyak I., and Graeser A. (2007) Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces, IEEE TBE 54(4): 742-750
8. Kay S. (1988) Modern Spectral Estimation: Theory and Application. Upper Saddle River, NJ: Prentice-Hall