MULTICHANNEL DECODING FOR PHASE-CODED SSVEP BRAIN-COMPUTER INTERFACE

NIKOLAY V. MANYAKOV∗, NIKOLAY CHUMERIN∗, MARC M. VAN HULLE
Laboratorium voor Neuro- en Psychofysiologie, KU Leuven,
Campus Gasthuisberg, O&N 2, Herestraat 49, 3000 Leuven, Belgium
E-mail: {NikolayV.Manyakov, Nikolay.Chumerin, Marc.VanHulle}@med.kuleuven.be

Received (to be inserted
Revised by Publisher)

We propose a complex-valued multilayer feedforward neural network classifier for decoding of phase-coded information from steady-state visual evoked potentials. To optimize the performance of the classifier we supply it with two filter-based feature selection strategies. The proposed approaches could be used for a phase-coded brain-computer interface, enabling to encode several targets using only one stimulation frequency. The proposed classifier is a multichannel one, which distinguishes our approach from the existing single-channel ones. We show that the proposed approach outperforms others in terms of accuracy and length of the data segments used for decoding. We show that the decoding based on one optimally selected channel yields an inferior performance compared to the one based on several features, which supports our argument for a multichannel approach.

1. Introduction

A Brain-Computer Interface (BCI) is a system that records and decodes brain activity so as to enable subjects to interact with the world through computers, robot actuators, and so on, bypassing the need for muscular activity. BCIs can significantly improve the quality of life for patients suffering from severe motor and/or communicative disabilities1,2. Brain-computer interfaces are either invasive3,4,5,6 or noninvasive7,8,9,10. In this paper, we consider a non-invasive BCI based on the steady-state visual evoked potential (SSVEP). It relies on the psychophysiological properties of EEG responses recorded from the occipital pole when observing repetitive (e.g., flickering) visual stimuli. Given that the periodic presentation is at a sufficiently high rate, the individual transient visual responses that are time and phase locked to the stimulus onset, will overlap and generate a steady state signal: the signal resonates at the stimulus rate and its multipliers11.

∗Equally contributed authors

Conventional SSVEP-based BCIs12,13,14,15 rely on the detection of an increase in amplitude at the frequencies f, 2f, 3f,... of the EEG signal’s power spectral density to infer that the subject is attending to a target flickering at rate f. Since the relevant EEG activity is always embedded in other ongoing brain activity and contaminated by (recording) noise, the detection task is not straightforward. To overcome this problem and to improve the decoding performance, several methods have been proposed: averaging over several time intervals12, recording over a longer period of time13, preliminary training14, etc. To enhance the BCI’s usability, several selectable targets (encoded by several frequencies) could to be used. This complicates the decoding process as one out of several frequencies needs to be selected by processing the EEG data.

Albeit that these methods were shown to achieve a reasonable information transfer rate15, the visual stimulation paradigm is facing a number of limi-
tations: only stimulus frequencies within a particular and subject-dependent frequency range evoke reasonable SSVEP responses\(^{14, 18}\); the harmonics of some stimulus frequencies could come close to other stimulus frequencies and their harmonics which eventually will affect the decoding performance\(^{16}\); when using a computer screen the stimulus frequency is restricted by the refresh rate \(f_{\text{scr}}\). These restrictions limit the number of targets in an SSVEP-based BCI. To increase the number of targets, the phase in the SSVEP response has been proposed in Ref.\(^{17, 18, 19}\), even a single frequency \(f\) could be used to encode \(N = f_{\text{scr}}/f\) commands by using the phase lag. This phase lag is produced by shifting in time the stimulus intensity profile by \(m\) frames (see Fig. 1), leading to a phase-shift of \(\Delta \phi_m = 2\pi m/N, m = 1, \ldots, N\). The targets constructed by the phase-lag are circular, since the shift of \(m\) frames in one direction is equivalent to one of \((N - m)\) frames in the opposite direction. Given that under normal conditions the individual SSVEP’s latency is stable\(^{18, 20}\), the delay in the stimulation would introduce the same delay in the recorded EEG data.

Most phase-coded SSVEP BCI systems use only a single channel (either Oz referenced to the mastoid\(^{17}\), or a bipolar lead\(^{18}\)) as the input to the decoder. Even though in Ref.\(^{21}\) a canonical correlation analysis (CCA) was proposed for a spatial filter fusing several channels into a single mixture-channel, which was further processed by a single-channel decoder. In this study, we present a classifier built on top of a multilayer neural network consisting of multi-valued neurons (MLMVN)\(^{22}\). The MLMVM network was originally designed to work with circular data and, therefore, it is in line with the considered classification problem.

\[
\begin{align*}
  f &= 10 \text{ Hz}, \quad \Delta \phi_1 = 0 & \ldots \\
  f &= 10 \text{ Hz}, \quad \Delta \phi_2 = \frac{\pi}{3} & \ldots \\
  f &= 10 \text{ Hz}, \quad \Delta \phi_3 = \frac{2\pi}{3} & \ldots \\
  f &= 10 \text{ Hz}, \quad \Delta \phi_4 = \pi & \ldots \\
  f &= 10 \text{ Hz}, \quad \Delta \phi_5 = \frac{4\pi}{3} & \ldots \\
  f &= 10 \text{ Hz}, \quad \Delta \phi_6 = \frac{5\pi}{3} & \ldots \\
\end{align*}
\]

Fig. 1. Phase-coded stimulation profiles. White squares indicate intensified frames, while dark squares indicate non-intensified frames. The screen refresh rate is 60 Hz; 18 video frames of stimulation are shown. Each row corresponds to one target stimulation profile.

We show the benefits of the multichannel approach by evaluating the dependency of the decoding accuracy on the number of input channels. To reduce the amount of irrelevant information to be processed by the classifier, we propose two filter-based feature selection techniques which rely on circular statistics to select the relevant channels.

2. Methods

2.1. EEG Data Acquisition

Our EEG recordings were made with a wireless device, developed by the Holst Centre\(^{23}\). Each EEG channel was sampled at 1024 Hz using 12 bit per sample. We used an EEG-cap with large filling holes and sockets for active Ag/AgCl electrodes (ActiCap, Brain Products). The recordings were made with eight electrodes located primarily on the occipital pole, namely at positions PO7, PO3, POz, PO4, PO8, O1, Oz, O2 according to the international 10–10 system. The reference and ground electrodes were placed on the right and left mastoids respectively. For further analysis, we additionally considered (eight) EEG signals from the mentioned electrodes measured with respect to common average reference (CAR) and all possible bipolar combinations, thus, \(D = 8 + 8 + C^2_8 = 44\) channels \(s_d(t)\).

The phases \(\varphi_d\) were estimated as \(\arg\left(\sum_i s_d(t) \cos(2\pi nft) + i \sum_i s_d(t) \sin(2\pi nft)\right)\), where \(i = \sqrt{-1}\), \(f\) is the stimulus frequency and \(n\) indicates the considered (sub)harmonic(s). We used segments \(s_d(t)\) of length \(T\) (\(T = 1, \ldots, 5\) seconds) cropped from the recordings starting from the stimulation onsets. Only the fundamental stimulus frequency was considered, thus \(n \equiv 1\), leading to \(D = 44\)-dimensional space of phases \(\varphi_d\).

2.2. Experiment description

Seven subjects (all male, aged 23–35, average 28.3 years) participated in the experiment. The subjects were sitting about 60 cm from the notebook’s LCD screen with reported refresh rate \(f_{\text{scr}} = 60\) Hz on which the stimuli of size 6 × 6 cm were shown. A set of \(N = 6\) stimuli flickering at \(f = 10\) Hz with phase shifts of \(\Delta \phi = \pi/3\) were simultaneously presented using the stimulation profile shown in Fig. 1. Each stimulus had a 50% duty cycle as this was reported to produce better detectable SSVEP re-
sponses for mostly all frequencies and for $f = 10$ Hz in particular\textsuperscript{24}. The stimuli were arranged in two rows and three columns, separated 7.5 cm horizontally and 7.75 cm vertically.

Each experiment consisted of 20 blocks. Each block was a sequence of six five-second long stimulation stages interleaved with one-second long “no-stimulation” stages, needed to shift the subject’s gaze to the next stimulus. During each stimulation stage the fixation point marker was placed on a flickering stimulus, thus the subject had to sequentially attend all six stimuli in one block.

2.3. Feature selection

In order to reduce the amount of information for subsequent classification, we propose a filter-based feature selection procedure. It selects only the most relevant features among all $D$ considered ones (the phases $\varphi_d$ extracted from all $D$ channels) for the classifier. For the same EEG data used in Fig. 2, the phases from the POz–Oz bipolar channel shown in panel (d) have a lower in-class scattering and demonstrate a better class separability compared to the phases estimated from the Oz channel (a), the POz–Oz bipolar channel (b), and when obtained after CCA spatial filtering (c). This suggests that the relevance of the features can be related to the data in-class scattering which, in turn, can be estimated by the standard deviation.

Given the circular nature of the data, we suggest to employ circular statistics\textsuperscript{25} to estimate the in-class standard deviation for each feature. In the first step the class means are estimated. For a set of phases $\Phi_d = \{\varphi_1^d, \varphi_2^d, \ldots, \varphi_L^d\}$ estimated for a channel $d$ in $L$ trials, the circular mean $\mu(\Phi_d)$ is estimated as $\arg z_d$, where $z_d = \sum_{l=1}^{L} \exp(i\varphi_{l}^d)$ and $i$ is the imaginary unit. The directions of the mean values for the data of each class are depicted by the radial lines in Fig. 2. The circular standard deviation of set $\Phi_d$ can be defined as $\sigma(\Phi_d) = \sqrt{2(1 - |z_d|)}$. Let $\sigma_d^m$ be the circular standard deviation of the phases estimated from the $d$-th channel for $m$-th class, then the value $\sigma_d = \sum_m \sigma_d^m$ can be considered as the cumulative scattering of channel $d$. Following the above mentioned observation (the lower $\sigma_d$, the “more relevant” channel $d$), the desired feature selection can be defined as the selection of the first $d$ channels sorted in ascending order according to their cumulative scattering $\sigma_d$. As a drawback, we can mention that the proposed technique does not take into account the between-class differences and, therefore, might select features that minimally contribute to the class separability.

Another, more reliable, feature selection method relies on a statistical test for differences between the mean values of the class pairs. Let us assume that the phases $\varphi_d$ from the $m$-th class, estimated from the $d$-th channel, are sampled from a von Mises distribution $p^m_d(\varphi|\mu^m_d, \kappa^m_d) = \exp(\kappa^m_d \cdot \cos(\varphi - \mu^m_d))/(2\pi I_0(\kappa^m_d))$, where $I_0$ is the modified zeroth-order Bessel function, and $\kappa^m_d$ and $\mu^m_d$ are the concentration and circular mean parameters. Given this assumption and the equalities of the $\kappa$’s, we can apply the pairwise (each class vs. each class, for every channel) Watson-Williams tests. After assigning to each channel the maximal $p$-value among all pairwise tests, one can re-order the channels to obtain the assigned values sorted in ascending order. The feature selection is then performed by taking the first $d$ first re-ordered channels.

3. Decoding based on MLNVN for phase-coded SSVEP BCI

Networks based on complex-valued neurons were reported to learn faster and to generalize better than traditional ones for different benchmarks and real world problems\textsuperscript{26,27}. We have used a multilayer feedforward neural network based on multi-valued neurons (MLMVN)\textsuperscript{22,27}. Such a network uses derivative-free backpropagation training, resulting in fast convergence to minimal error rates\textsuperscript{22}. In MLMVN, each $k$-th neuron from every hidden or output layer $(j$-th layer) has connections to all neurons from the previous ($(j - 1)$-th) layer and has a complex activation function leading to the output $y_{k,j} = z_{k,j}/|z_{k,j}|$, where $z_{k,j} = w_{0}^{k,j} + w_{1}^{k,j} y_{1,j-1} + \cdots + w_{N_{j-1}}^{k,j} y_{N_{j-1},j-1} = (\overline{w}_j^k)^T \overline{y}_{j-1}$, $y_{1,j-1} \in \mathbb{C}$, $|y_{1,j-1}| = 1$ is the output of the $l$-th neuron from the previous $(j - 1)$-th layer, $w_{k,j}^{1,j} \in \mathbb{C}$ is the corresponding weight connecting the $l$-th neuron to the $k$-th neuron in the next $j$-th layer, $\overline{w}_j^k = (w_{0}^{k,j}, \ldots, w_{N_{j-1}}^{k,j})^T$ and $\overline{y}_{j-1} = (1, y_{1,j-1}, \ldots, y_{N_{j-1},j-1})^T$ and $N_{j-1}$ is the number of neurons in the $(j - 1)$-th layer.

Let us consider a MLMVN comprising $(n - 1)$ hidden layers and one output layer with a single
neuron. The network’s global error is estimated as \( E = \theta - y_{1,n} \), where \( \theta \) is the desired output and \( y_{1,n} \) is the network’s actual output. Assuming that the redistribution of this error between all neurons and given the threshold in the previous layer’s neurons, we get \( \delta_{1,n} = \frac{1}{N_{j-1}+1} E \). This error is then backpropagated according to \( \delta_{kj} = \frac{1}{N_{j-1}+1} \sum_{i=1}^{N_{j-1}+1} \delta_{1,s} (w_{k,j}^{(i+j+1)})^{-1} \). During training, the weights are adjusted as \( w_{lj}^k = w_{lj}^k + \frac{1}{(N_{j-1}+1)C_{kj}} \tilde{\delta}_{l}^T Y_{j-1} \), where \( \tilde{\delta}_{l} = (\delta_{lj}, \ldots, \delta_{N_{j}lj})^T \) and \( C_{kj} \) is equal to 1 for the last layer and \( C_{kj} = |z_{k,j}| \) for all other cases.27

It is convenient to represent the circular input and output data as complex numbers of unit length. Thus, all the selected \( d \) phase values \( \varphi_{d} \) from the preselected channels can be represented by complex numbers \( \exp(i\varphi_{d}) \), which can be further used as inputs to the MLMVN.

During training, for the training samples of class \( m \) we used as the desired network’s output value \( \theta_{m} = \exp(i2\pi(m - \frac{1}{2})/N) \). Then, from the output \( y_{1,n} \) of the network, the resulting class index \( \tilde{m} \) is deduced as an integer satisfying two conditions: \( 2\pi(\tilde{m} - 1)/N \leq \arg y_{1,n} < 2\pi\tilde{m}/N \) and \( 1 \leq \tilde{m} \leq N \).

During training we kept track of an angular variant of the root mean square error (RMSE) and stopped the training, when the RMSE became lower than 0.1 radian.

For our experiments, we used an MLMVN with a single hidden layer. This choice was motivated by the next observation. The use of a single multi-valued neuron for our problem did not allow us to get a proper class separability since training did not decrease the training error below the chosen threshold due to the more complex nature of the separation problem. This calls for an MLMVN. Since we did not observe any significant improvement in performance by increasing the number of hidden layers, but only by increasing the training time, we decided to stick to one hidden layer.

4. Results

Prior to the actual experiment, we had to select the architecture of the MLMVN, namely, the number of neurons in the hidden layer \( N_h \) and the number of features required for obtaining a satisfactory decoding accuracy (i.e., the network input dimensionality). To set these parameters up, we collected data from two subjects, and assessed the classification accuracy based on a five-fold cross-validation for all combinations of \( N_h = 2, \ldots, 20 \) and \( d = 1, \ldots, 44 \) with both proposed feature selection methods. Based on the results (not shown), we considered as the MLMVN’s input the four best separating features according to our heuristic based on the standard deviation (since both feature selection methods perform almost equally well in terms of accuracy) and \( N_h = 10 \). The incorporation of more features decreased the classification performance. The reason for this we see in forcing the classifier to deal with irrelevant information, which consequently deteriorates the generalization performance of the neural network. The selected \( N_h = 10 \) can be seen as a trade-off between training time and flexibility of the network (with large \( N_h \)), which could cause overfitting and reduce the generalization performance.

Figure 2: Distribution of phases (expressed as angular values) estimated from the same experimental (see Sec. 2.2) data for subject 1 from Oz (a), POz–Oz (b), after CCA spatial filtering (c) and POz–O2 (d). Each dot corresponds to a phase estimated from a one second long interval recorded when the subject was observing a particular phase-shifted stimulus. Colors represent target-classes (with the stimulus shifted by \( \Delta \phi_{m} = m\pi/3 \), with \( m \) the class index). Radial lines correspond to circular means for each class. For the sake of visualization, each class is drawn on a circle with a different radius.
Figure 3 shows the result of a five-fold cross-validation performed on the methods of Ref.\textsuperscript{17,18}, method based on CCA spatial filtering\textsuperscript{21} and the proposed MLMVN-based method, for different EEG interval lengths $T$ used for phase estimation. It shows that the results obtained with the MLMVN significantly outperform the other considered methods according to the Wilcoxon signed rank test ($p < 0.05$), at least for the tested subjects. By applying the single-channel methods\textsuperscript{18,17} to the optimally selected (via a wrapper-like exhaustive search through all channels $s_d$ on the training data) channel, we also observe the superiority of the proposed multichannel classifier as indicated by the $p$-values in Fig. 3.

### 5. Discussion

In Ref.\textsuperscript{18} the case of one optimal channel for classification was considered, whereas we proposed a classifier that incorporates several channels which produces better results. Hence, the question arises whether actually the consideration of several channels could improve the proposed classifier performance? As one can see from Fig. 4, the averaged classification accuracy increases up to a maximum when considering features from about 3–5 “best” channels. A statistical comparison of the accuracy achieved with a single channel and with the optimal number of channels reveals a difference with $p$-value (of the Wilcoxon signed rank test) decreases from 0.25 to 0.03 with decreasing $T$. This supports our hypothesis that the proposed multichannel classifier outperforms considered the single-channel classifiers.

In Ref.\textsuperscript{18} the dependence of the decoding performance on the number of harmonics was analyzed. It turned out that the inclusion of the additional harmonic(s) improves the classification accuracy. In this study we do not consider harmonics and use only the fundamental frequency ($n = 1$). But we have to say that the proposed decoding algorithm is capable of incorporating any additional circular features as, for example, the phases estimated from the harmonics of the considered channels. This might lead to better results. The proposed filter-based feature selection methods process each channel separately, thus, not jointly. Moreover, they are quite sensitive to outliers especially on a small training set. One of the possible solutions could rely on a wrapper-based selection procedure involving the proposed MLMVN classifier.

The solution of the above mentioned issues we see as future steps in phase-coded SSVEP-based BCI research. These steps also include the on-line assessment of the proposed system from the point of view of human-computer interaction. As was hypothesized in Ref.\textsuperscript{21}, neighboring flickering stimuli influence the estimated phases. To reduce this influence and increase the delectability one can employ a special layout of the stimuli displayed on the screen. Another direction is the design of a new stimulation profile, with which one can involve arbitrary phase shift $\Delta \phi_m$. Finally, the proposed method can be extended to the case of several stimulation frequencies fusing the conventional frequency-based SSVEP with the phase-based SSVEP methods for BCIs.

![Figure 3: Average discrimination accuracy as a function of the EEG segment length](image)

![Figure 4: Averaged subject accuracy plotted as a function of the number of best features selected by the method based on the standard deviation (std) and for $N_h = 10$ and different segment lengths $T$.](image)

### Acknowledgments

NVM is supported by the research grant GOA 10/019, NC is supported by Tetra project Spellbinder, MMVH is supported by PFV/10/008, CREA/07/027, G.0588.09, IUAP P6/29, GOA 10/019 and the Tetra project Spellbinder.
Multichannel phase-coded SSVEP-based BCI

References