

Advancing the detection of steady-state visual evoked potentials in brain-computer interfaces

Mohammad Abu-Alqumsan¹ and Angelika Peer²

¹ Chair of Automatic Control Engineering, Technical University of Munich (TUM), Munich, Germany

² Bristol Robotics Laboratory, University of the West of England, Bristol, UK

E-mail: moh.marwan@lsr.ei.tum.de

Abstract. *Objective.* Spatial filtering has proved to be a powerful pre-processing step in detection of steady-state visual evoked potentials and boosted typical detection rates both in offline analysis and online SSVEP-based brain-computer interface applications. State-of-the-art detection methods and the spatial filters used thereby share many common foundations as they all build upon the second order statistics of the acquired EEG data, that is, its spatial autocovariance and cross-covariance with what is assumed to be a pure SSVEP response. The present study aims at highlighting the similarities and differences between these methods. *Approach.* We consider the canonical correlation analysis (CCA) method as a basis for the theoretical and empirical (with real EEG data) analysis of the state-of-the-art detection methods and the spatial filters used thereby. We build upon the findings of this analysis and prior research and propose a new detection method (CVARS) that combines the power of the canonical variates and that of the autoregressive spectral analysis in estimating the signal and noise power levels. *Main results.* We found that the multivariate synchronization index (MSI) method and the maximum contrast combination (MCC) method are variations of the CCA method. All these three methods were found to provide relatively unreliable detections in low SNR regimes. CVARS and the minimum energy combination (MEC) methods were found to provide better estimates for different SNR levels. *Significance.* Our theoretical and empirical results demonstrate that the proposed CVARS method outperforms other state-of-the-art detection methods when used in an unsupervised fashion. Furthermore, when used in a supervised fashion, a linear classifier learned from a short training session is able to estimate the hidden user intention, including the idle state (when the user is not attending to any stimulus), rapidly, accurately and reliably.

Keywords: BCI, SSVEP, canonical correlation analysis, minimum energy combination, multivariate synchronization index, maximum contrast combination, autoregressive spectral analysis

Submitted to: *J. Neural Eng.*

1. Introduction and Motivation

Steady-state visual evoked potentials (SSVEP) refer to the involuntary brain response to repetitive visual stimulation of the eye in humans and some non-human primates [1]. The measured scalp Electroencephalographic (EEG) potentials are characterized in the frequency domain by constant amplitude and phase at the Fourier components of the stimulation frequency and its higher harmonics. SSVEPs can be observed when the driving frequency of the stimulus is in the range 4 – 100 Hz [2], below which the electrical excitations of the visual system is able to abate before the new stimuli are presented, and in this case, it is the transient VEPs that can be observed [3]. Capilla et al. [4] showed that steady-state VEPs can be accurately predicted from the linear summation of appropriately constructed transient responses and concluded that both can be attributed to the same underlying neural mechanism.

Due to several factors, SSVEP has been a core concept in non-invasive EEG-based brain-computer interface (BCI) applications. Among these are its robustness and relatively high signal-to-noise ratio (SNR), the high information transfer rate (ITR) it delivers and the short training time required, if any, before it can be used for online applications. Typically, in SSVEP-based BCIs, different frequency-tagged stimuli are displayed simultaneously with each stimulus given a predefined mapping to a system command. This mapping is also known to the user, who can control or communicate with the system, simply by attending to the stimulus corresponding to the command of interest. Relatively easy discrimination between the different frequencies is facilitated by the fact that selective attention to stimulus location modulates SSVEP [5].

Early investigated methods for SSVEP analysis and detection have relied mainly on Power Spectral Density Analysis (PSDA) with Fast Fourier Transform (FFT) applied on single-channel EEG data. Herewith, the temporal EEG signal is transformed into its Fourier representations where test statistics can be derived from obtained information about the power (or amplitude) [3, 6, 7, 8, 9], phase [10] or both [11], at all considered driving frequencies. FFT-based methods require relatively long data segments to give reasonable detection results since the frequency resolution (Δf) in the Fourier domain is determined by the reciprocal of

the temporal data length available (e.g. it is required to have a 4 s data segment to get a frequency resolution of $\Delta f = 0.25$ Hz). Known issues with FFT like the grid effect (Fourier components cannot be computed for frequencies that are not an integral multiple of Δf) and spectral leakage (energy spillover from one frequency bin to adjacent ones due to rectangular windowing) highly affect the calculated amplitude and phase precision and should be accounted for by choosing suitable window functions and segment lengths [5, 12, 13]. In [14], authors argue that for an arbitrary steady-state recording, there might not be enough information to benefit from phase information in SSVEP detection.

Recently, there has been a great tendency towards different methods that rely on spatial filtering of multi-channel EEG data, which proved to be more efficient and stable than FFT-based methods. The basic idea here is to find a spatial filter that transforms the original multi-channel EEG signals into single or multi-channels with desirable characteristics. Friman et al. [15] proposed the minimum energy combination (MEC) spatial filter, the first of its kind, which aims at minimizing the noise energy, with the noise defined as the orthogonal complement to the projection of the EEG signals onto the subspace spanned by the pure SSVEP vectors. Alternatively, the canonical correlation analysis (CCA) aims at finding a pair of spatial linear combinations for both, the EEG signals and the assumed pure SSVEP responses, which jointly maximize the correlation between the resulting canonical variates [16]. The maximum contrast combination (MCC) filter maximizes the signal-to-noise ratio, defined with the generalized Rayleigh quotient [15, 17]. Finally, the multivariate synchronization index (MSI), through spatial whitening, extracts a single metric that reflects the synchronization level between the EEG signals and assumed pure SSVEPs [18].

The main contribution of this work is that we show the similarity of the spatial filters and the scoring functions used by the standard unsupervised CCA, MSI and MCC approaches to SSVEP detection, where we conclude also that their detection accuracies should not differ significantly. We show additionally that these methods, in low signal-to-noise ratio (SNR) regimes, fail to provide a reliable detection as they ignore the noise power in the spontaneous EEG that overlaps the stimulation frequencies.

On the other hand, parametric spectral analysis,

like autoregressive methods, are able to provide an estimate for the noise power levels after removing the total energy of the driving frequencies from the spatially filtered signals. This was shown to provide reliable test statistics for SNR that can be used for SSVEP detection [15, 19]. Based on the theoretical analysis of these state-of-the-art methods, we propose a new method, namely the *canonical variates with autoregressive spectral analysis* (CVARS) that estimates the signal and noise power levels from the canonical variates which leads to slight improvement in correct detection rates. This method, when used in a supervised fashion, results in a rapid, accurate and reliable SSVEP detection regardless of the level of SNR as shown from the experimental results with real EEG data. We additionally show the conditions on which the MEC-based detection provides similar results to the CVARS method.

2. Preliminaries

2.1. Notation

We use bold uppercase letters (e.g. \mathbf{Z}) to denote two-dimensional matrices and bold lowercase letters (e.g. \mathbf{z}) to denote column vectors, and conversely \mathbf{z}^T for row vectors. The Frobenius norm of the matrix \mathbf{Z} is denoted by $\|\mathbf{Z}\|_F$.

A p -variate random variable is denoted by an uppercase letter, where it refers to a vector of p random variables, e.g. $Z = [Z_1, Z_2, \dots, Z_p]$. Realizations of Z are denoted the same way as vectors (e.g. \mathbf{z}). A sequence of n p -dimensional vectors is concatenated in the sample data matrix $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n]^T$, where $\mathbf{Z} \in \mathbb{R}^{n \times p}$. The expectation of a random variable Z is denoted by $E[z]$ and $E[\mathbf{z}]$, for one and multi-dimensional variables, respectively. For an arbitrary matrix $\mathbf{Z} \in \mathbb{R}^{n \times p}$, $\sigma(\mathbf{Z}) = (\sigma_1, \sigma_2, \dots, \sigma_{\min(n,p)})$ denotes the singular values of \mathbf{Z} , where $\sigma_1 \geq \sigma_2 \geq \dots \sigma_{\min(n,p)} \geq 0$. For arbitrary square matrix $\mathbf{Z} \in \mathbb{R}^{n \times n}$, $\lambda(\mathbf{Z}) = (\lambda_1, \lambda_2, \dots, \lambda_n)$ denotes the eigenvalues of \mathbf{Z} with $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n$. The dominant eigenvalue is denoted by $\lambda_1(\mathbf{Z})$ or $\lambda_{\max}(\mathbf{Z})$. An arbitrary diagonal matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ can be defined with its diagonal elements as $\text{diag}(d_1, \dots, d_n)$.

The autocovariance matrix of p -variate random variable X is denoted by \mathbf{C}_{xx} and the cross-covariance matrix between two random vectors X and Y is denoted by \mathbf{C}_{xy} . Often in practice, the true value of such covariances remains unknown, and therefore they are replaced by the sample auto and cross-covariance matrices \mathbf{Q}_{xx} and \mathbf{Q}_{xy} , computed from the centered sample matrices with $\mathbf{Q}_{xx} = \frac{1}{N} \mathbf{X}^T \mathbf{X}$ and $\mathbf{Q}_{xy} = \frac{1}{N} \mathbf{X}^T \mathbf{Y}$, where N is the number of available samples in \mathbf{X} and \mathbf{Y} . Unless explicitly stated otherwise, all sample matrices encountered here will be centered or

made centered by subtracting the sample mean from all observations.

2.2. Background

We assume that after the retina is excited by flickering light, and after the transient VEPs vanish [20], pure SSVEP responses appear as multiple phase-shifted sinusoidal waves whose frequencies are integer multiple (up to N_h) of the driving frequency. The assumed pure waves propagate to the scalp where EEG signals are measured. Due to volume conduction of the head, a linear combination of these source signals corrupted with noise will be observed at each measurement location, i.e. electrode. The additive noise might have encephalic or non-encephalic sources, and is generally non-stationary. However, for short EEG segments, it is often assumed to be stationary in the wide sense [21]. No statistical knowledge about the noise is additionally assumed here. Formally, we state that in response to flickering light with driving frequency f , the values recorded over time at each electrode i can be written as

$$y_i(t) = \sum_{h=1}^{N_h} a_{i,h} \sin(2\pi h f t + \phi_{i,h}) + \psi_i(t), \quad (1)$$

where $i \in \{1, \dots, N_y\}$ and N_y is the total number of recording electrodes. The terms $a_{i,h}$ and $\phi_{i,h} \in]-\pi, \pi]$ respectively denote the harmonic-specific amplitude and phase lag at electrode i . The additive noise contributing to $y_i(t)$ is denoted by $\psi_i(t)$. For the digitized EEG with sampling rate of F_s (in Hz), (1) becomes

$$y_i[k] = \sum_{h=1}^{N_h} a_{i,h} x_{i,h}[k] + \psi_i[k], \quad \text{where} \quad (2)$$

$$\begin{aligned} x_{i,h}[k] &= \sin(2\pi h f \frac{k}{F_s} + \phi_{i,h}) \\ &= \sin(2\pi h f \frac{k}{F_s}) \cos(\phi_{i,h}) + \cos(2\pi h f \frac{k}{F_s}) \sin(\phi_{i,h}). \end{aligned}$$

With conformable transformation of the amplitude values $a_{i,h}$, the above notation can be rewritten for a collection of N samples from all N_y electrodes (see figure 1) in matrix form as

$$\mathbf{Y} = \mathbf{X} \mathbf{A} + \mathbf{\Psi}. \quad (3)$$

Here, $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_{N_y}]$, $\mathbf{\Psi} \in \mathbb{R}^{N \times N_y}$ and the propagation (or mixing) matrix $\mathbf{A} \in \mathbb{R}^{2N_h \times N_y}$. The source model matrix $\mathbf{X} \in \mathbb{R}^{N \times 2N_h}$ is defined now as

$$\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_h \ \dots \ \mathbf{X}_{N_h}], \quad (4)$$

where $\mathbf{X}_h = [\sin(2\pi h f k') \ \cos(2\pi h f k')] \in \mathbb{R}^{N \times 2}$, and $k' = \frac{1}{F_s}, \frac{2}{F_s}, \dots, \frac{N}{F_s}$.

In absence of any statistical knowledge about the noise, \mathbf{A} can be estimated from the source model and

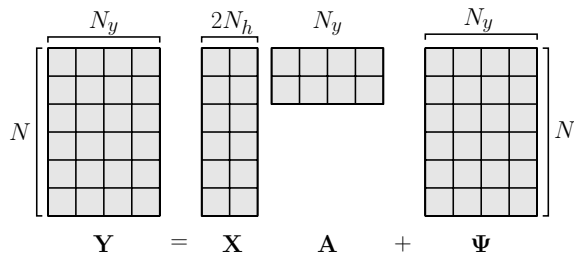


Figure 1: Matrix representation of the EEG model containing SSVEPs

the acquired EEG data by $\mathbf{A}_{LS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ which minimizes the total least squares error, i.e. $\|\mathbf{Y} - \mathbf{X}\mathbf{A}\|_F^2$. Equivalently, the LS estimate of \mathbf{A}_{LS} can be written as $\mathbf{A}_{LS} = \mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy}$, where \mathbf{Q}_{xx} and \mathbf{Q}_{xy} respectively denote the sample spatial autocovariance matrix of the pure SSVEP response and the sample cross-covariance between the acquired EEG data and the assumed source model, both are estimated from N observations. Despite the fact that the source signals are defined deterministically, they can be modeled as stochastic with empirical means computed on segments of length N , which can be approximated with the zero-vector $\mathbf{0}$ for large N . Due to the orthogonality of the basis vectors in \mathbf{X} , the sample covariance matrix \mathbf{Q}_{xx} , for centered and normalized \mathbf{X} , can also be approximated for relatively large N with \mathbf{I}_{N_x} , where $N_x = 2N_h$. In this formulation, the number of harmonics in the SSVEP response, i.e. N_h , is not a random variable but rather an unknown deterministic value.

Where it is necessary, in order to avoid ambiguity when we refer to the source model for the different driving frequencies, the subscript f will be added to matrix \mathbf{X}_f to indicate the source model of the driving frequency under consideration.

2.3. Stimulus identification during SSVEP-based interaction

During concurrent repetitive visual stimulation, an SSVEP detector aims at finding the stimulus, at which the user is attending, based on multi-channel EEG data segment $\mathbf{Y} \in \mathbb{R}^{N \times N_y}$, obtained online from continuous scalp EEG data by means of buffering (with buffer length N and buffer overlap O where $0 \leq O < N$). Based on the requirements of the application at hand, the buffer length and the interaction temporal resolution (T) are determined. T is defined as the shortest time (in samples) between two consecutive commands. The overlap can be computed from $O = \max(0, N - T)$.

The problem of stimulus identification from EEG data thus can be formulated as having M spatially-

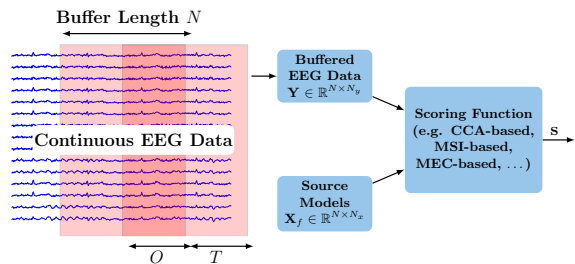


Figure 2: Schematic of a general SSVEP detector from continuous EEG data. Different scoring functions can be used to provide the score vector $\mathbf{s} \in \mathbb{R}^{M+1}$, whereby a decision about the user intention can be made every T samples.

distributed flickering lights, driven by different frequencies f_1, f_2, \dots, f_M , and a mapping function \hat{g} is sought, where $\hat{g} : \mathbb{R}^{N \times N_y} \mapsto \{f_0, f_1, \dots, f_M\}$, and f_0 denotes the idle state, i.e. when the user does not attend to any of the stimuli. Often, $\hat{g}(\mathbf{Y})$ is defined as the argument which maximizes a score function or a test statistic s . SSVEP detection can be formally written as

$$\hat{f} = \hat{g}(\mathbf{Y}) = \underset{f_i \in \{f_0, f_1, \dots, f_M\}}{\operatorname{argmax}} s(\mathbf{Y}, \mathbf{X}_{f_i}). \quad (5)$$

We denote by $g(\mathbf{Y})$ the ground truth frequency of the stimulus, to which the user attends while \mathbf{Y} is being acquired. The score $s(\mathbf{Y}, \mathbf{X}_{f_0})$ is considered here as to test whether or not a given response is statistically significant and not due to noise fluctuations and background EEG. More often than not, it is defined as a constant threshold which is either computed from the EEG data segments themselves or a priori computed from training data. For completeness, let $\mathbf{s} \in \mathbb{R}^{(M+1) \times 1}$ denote the vector containing the score value for all frequencies plus the idle state.

Figure 2 depicts the schematic of SSVEP detection in continuous EEG data. Available scoring functions used in (5) will be discussed in section 3.

2.4. Evaluation

The different detection methods, which will be discussed later, are compared with regard to their average accuracy (\bar{P}_D), which is typically computed from labeled EEG segments as the ratio of the correctly classified segments to the total number of available segments. Average misclassification error can be easily computed with $E_m = 1 - \bar{P}_D$.

2.5. Challenges in SSVEP-based interaction

The user in SSVEP-mediated applications continuously shifts his/her gaze between the M spatially distributed stimuli for active control and towards the stimuli-free areas of the display for the idle control

state. Obviously, such interaction is asynchronous and completely paced with the user actions. However, and regardless of the SSVEP detection method used, recognition of these asynchronous spatial attention shifts does not happen usually on the spot due to inherent limiting factors of the buffering stage. The shorter the buffer size (small N and T), the faster is the response of the system. With larger buffer sizes, however, the temporal random fluctuations in the score function $s(\mathbf{Y}, \mathbf{X}_f)$ are made less severe as the noise (the assumed source of variability in the evoked potentials) attenuates typically in proportion to the square root of the number of time averages done on the data. On the other hand, larger buffer sizes introduce delays into the system and reduce the achievable bit rate. Consequently, finding a trade-off between interaction accuracy and speed is of high importance for practical systems.

3. SSVEP detection methods

In the next subsections, we will provide the theoretical background for the spatial filtering methods used in state-of-the-art SSVEP detection and show in which ways they differ and how the spatial filters obtained are related to each other. Additionally, we propose a new detection method that performs superiorly in different SNR regimes as it combines CCA spatial filtering and autoregressive spectral analysis to estimate the signal and noise levels at all possible driving frequencies.

3.1. Canonical correlation Analysis (CCA)

Lin et al. [16] used canonical correlation analysis (CCA) to recognize the narrow-band driving frequency of SSVEPs from EEG data. The CCA-based method was found to outperform the FFT-based spectrum estimation method in terms of classification accuracy. This result has been repeatedly reported in [8, 22]. The superior performance can be attributed to the ability of CCA to reveal spatial coherence in data contaminated by either white Gaussian noise or colored noise fields, should the data have high SNR [23].

CCA [24] does that by finding the maximally correlated pairs among all possible linear combinations of two zero-mean multivariate random variables X and Y , where $\mathbf{x} \in \mathbb{R}^{N_x \times 1}$ and $\mathbf{y} \in \mathbb{R}^{N_y \times 1}$. Without loss of generality, we assume in the following that $N_x \leq N_y$.

Formally, we look for the canonical weight vectors \mathbf{w}_x and \mathbf{w}_y where $x = \mathbf{w}_x^T \mathbf{x}$ and $y = \mathbf{w}_y^T \mathbf{y}$, such that the correlation coefficient between the canonical

variates x and y , $\rho_1(x, y)$ is maximized. By definition,

$$\begin{aligned} \rho_1(x, y) &= \frac{E[xy]}{\sqrt{(E[x^2]E[y^2])}} \\ &= \frac{E[\mathbf{w}_x^T \mathbf{x} \mathbf{y}^T \mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T \mathbf{x} \mathbf{x}^T \mathbf{w}_x] E[\mathbf{w}_y^T \mathbf{y} \mathbf{y}^T \mathbf{w}_y]}} \\ &= \frac{\mathbf{w}_x^T E[\mathbf{x} \mathbf{y}^T] \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T E[\mathbf{x} \mathbf{x}^T] \mathbf{w}_x \mathbf{w}_y^T E[\mathbf{y} \mathbf{y}^T] \mathbf{w}_y}} \\ &= \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}. \end{aligned} \quad (6)$$

Since scaling of \mathbf{w}_x and \mathbf{w}_y doesn't affect the objective function, the search space is limited by constraining the variance of the variates x and y to be 1 [25]. This leads to the new optimization problem

$$\begin{aligned} \mathbf{w}_x, \mathbf{w}_y &= \underset{\mathbf{w}_x, \mathbf{w}_y}{\operatorname{argmax}} \mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y \\ &\text{subject to } \mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x = \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y = 1 \end{aligned}$$

By introducing Lagrange multipliers, one can easily obtain the following generalized eigenvalue problems

$$\mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{w}_x = \rho_1^2 \mathbf{C}_{xx} \mathbf{w}_x \quad (7)$$

$$\mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{w}_y = \rho_1^2 \mathbf{C}_{yy} \mathbf{w}_y. \quad (8)$$

Due to the fact that \mathbf{C}_{xx} and \mathbf{C}_{yy} denote covariance matrices, which are symmetric positive semi-definite, (7) and (8) can be rearranged into two standard symmetric eigenvalue problems,

$$\mathbf{T} \mathbf{T}^T \mathbf{w}'_x = \rho_1^2 \mathbf{w}'_x, \quad (9)$$

$$\mathbf{T}^T \mathbf{T} \mathbf{w}'_y = \rho_1^2 \mathbf{w}'_y, \quad (10)$$

where $\mathbf{T} = \mathbf{C}_{xx}^{-1/2} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1/2}$, $\mathbf{w}'_x = \mathbf{C}_{xx}^{1/2} \mathbf{w}_x$ and similarly $\mathbf{w}'_y = \mathbf{C}_{yy}^{1/2} \mathbf{w}_y$. The matrix \mathbf{T} is referred to as the *coherence matrix* and denotes the cross-covariance between the whitened vectors $\mathbf{C}_{xx}^{-1/2} \mathbf{x}$ and $\mathbf{C}_{yy}^{-1/2} \mathbf{y}$. This yields that $\rho_1(x, y) = \sigma_{\max}(\mathbf{T}) = \sqrt{\lambda_{\max}(\mathbf{T} \mathbf{T}^T)}$. The matrix product $\mathbf{T} \mathbf{T}^T$ is often referred to as *squared coherence matrix* [26]. Other uncorrelated canonical variates can be found using the remaining eigenvectors and eigenvalues [25].

For later use, we define the decomposition $\mathbf{T}^T \mathbf{T} = \mathbf{W}'_y \mathbf{P}^2 \mathbf{W}'_y{}^T$, where $\mathbf{W}'_y \in \mathbb{R}^{N_y \times N_x}$ and \mathbf{P} has all the canonical correlations on its diagonal, i.e. $\mathbf{P} = \operatorname{diag}(\rho_1, \rho_2, \dots, \rho_{N_x}) = \operatorname{diag}(\sqrt{\lambda(\mathbf{T} \mathbf{T}^T)})$. Similarly, $\mathbf{T} \mathbf{T}^T = \mathbf{W}'_x \mathbf{P}^2 \mathbf{W}'_x{}^T$, where $\mathbf{W}'_x \in \mathbb{R}^{N_x \times N_x}$.

The score function in (5) can thus be defined for the standard CCA method as $s_{\text{cca}} = \rho_1$. Alternatively, other score functions can be derived as an arbitrary function of $(\rho_1, \dots, \rho_{N_x})$, in the form $s_{\text{fcc}} = f(\sigma(\mathbf{T}))$.

Since true covariance matrices are not known a priori, the coherence matrix is defined with the empirical estimates thereof. Recall that $\mathbf{Q}_{xx} \approx \mathbf{I}_{N_x}$.

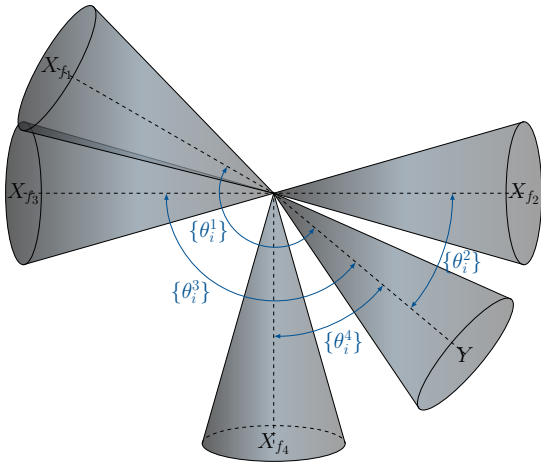


Figure 3: Principal angles between the subspaces spanned by the different source models and the acquired EEG data. Each is a subspace in \mathbb{R}^N and is represented by a convex cone. Note also that the column spaces of the different \mathbf{X}_{f_i} may intersect if they share, at least, one basis vector, as it is the case with \mathbf{X}_{f_1} and \mathbf{X}_{f_3} , where $f_1 = 15$ and $f_3 = 10$, and $N_h = 3$.

Given enough samples for \mathbf{Y} , i.e. $N \gg N_y$, then \mathbf{Q}_{yy} will be full rank and invertible.

Equivalently, the canonical correlations [27] can be found by first applying the QR decomposition of $\mathbf{Y} = \mathbf{Q}_y \mathbf{R}_y$ and $\mathbf{X} = \mathbf{Q}_x \mathbf{R}_x$, to obtain the orthonormal matrices \mathbf{Q}_x and \mathbf{Q}_y and the full rank \mathbf{R}_x and \mathbf{R}_y matrices. The second step involves the singular value decomposition of $\mathbf{Q}_x^T \mathbf{Q}_y$ as $\mathbf{U} \mathbf{P} \mathbf{V}^T$, where $\mathbf{U} = \frac{1}{\sqrt{N}} \mathbf{R}_x \mathbf{W}_x$ and $\mathbf{V} = \frac{1}{\sqrt{N}} \mathbf{R}_y \mathbf{W}_y$.

This formulation is computationally more efficient and provides more insights on the geometric interpretation of CCA. Hereby, the canonical correlations correspond to the cosine of the principal angles between the two subspaces spanned by the column spaces of \mathbf{Q}_x and \mathbf{Q}_y or formally, $\sigma(\mathbf{Q}_x^T \mathbf{Q}_y) = (\cos(\theta_1), \dots, \cos(\theta_{\min(N_y, N_x)}))$. Geometrically, the maximum canonical correlation is the cosine of the smallest angle possible between any two vectors in the subspaces spanned by \mathbf{Q}_x and \mathbf{Q}_y . Figure 3 illustrates this relation.

3.2. Multivariate Synchronization Index (MSI)

More recently, Zhang et al. [18] introduced the multivariate synchronization index (MSI) for online SSVEP detection, where the synchronization level between the source model and the acquired EEG is measured based on the S-estimator [28]. The joint covariance matrix $\mathbf{C}_{X,Y}$ which includes the auto and cross-covariance matrices of \mathbf{X} and \mathbf{Y} can be written

in block form as

$$\mathbf{C}_{X,Y} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}. \quad (11)$$

The transform \mathbf{U} that orthogonalizes the diagonal block matrices, i.e. whitens the original data matrices \mathbf{X} and \mathbf{Y} , was applied such that $\mathbf{R} = \mathbf{U} \mathbf{C}_{X,Y} \mathbf{U}^T$ and

$$\mathbf{U} = \begin{bmatrix} \mathbf{C}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{yy}^{-1/2} \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \mathbf{I}_{N_x} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{I}_{N_y} \end{bmatrix}, \quad (12)$$

where $\mathbf{T} = \mathbf{C}_{xx}^{-1/2} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1/2}$ is the coherence matrix already encountered. Let $P = 2N_h + N_y$, $\lambda(\mathbf{R}) = (\lambda_1, \dots, \lambda_P)$, and the normalized eigenvalues to be defined as $\lambda'_i = \frac{\lambda_i}{P}$. The synchronization index then can be obtained from the entropy-like quantity

$$s_{\text{msi}} = 1 + \frac{\sum_{i=1}^P \lambda'_i \log(\lambda'_i)}{\log(P)}. \quad (13)$$

The MSI-based score is tightly related to s_{cca} , as the eigenvalues of \mathbf{R} are nothing more than a simple transformation of the canonical correlations, such that $\lambda_i = 1 + \rho_i, \forall i \in \{1, \dots, N_x\}$, $\lambda_i = 1 - \rho_i, \forall i \in \{N_x + 1, \dots, 2N_x\}$ and $\lambda_i = 1$, otherwise (see Appendix A). This renders s_{msi} as a mere nonlinear function of all canonical correlations, and as special case of the s_{fcca} . Additionally, the filtering step (i.e. whitening) involved is similar to that in the CCA method.

3.3. Minimum Energy Combination (MEC)

Friman et al. [15] proposed to apply the spatial filter $\mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N_y \times N_s}$, which minimizes the noise energy in $\mathbf{S} = \mathbf{Y} \mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N \times N_s}$, where $N_s \leq N_y$. The noise here is defined as the difference between the original EEG signal and its best LS approximation in the subspace spanned by the SSVEP sinusoids. The noise is thus estimated as

$$\tilde{\Psi} = \mathbf{Y} - \mathbf{X} \mathbf{A}_{LS} = \mathbf{Y} - \mathbf{X} \mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy}. \quad (14)$$

Note that $s_{\text{cca}}(\tilde{\Psi}, \mathbf{X}) = 0$. The sample noise covariance matrix can be written as $\mathbf{Q}_{\psi\psi} = \mathbf{Q}_{yy} - \mathbf{Q}_{yx} \mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy}$ with the eigendecomposition $\mathbf{Q}_{\psi\psi} \mathbf{D}_{\psi} \mathbf{Q}_{\psi}^T$. The spatial filter \mathbf{W}_{MEC} is then obtained by concatenating the last N_s vectors in \mathbf{Q}_{ψ} which correspond to the least proportion of energy in $\tilde{\Psi}$, e.g. correspond to the eigenvalues whose sum does not exceed $\text{Tr}(\mathbf{Q}_{\psi\psi})/10$ [15].

A test statistic can be derived from the filtered signals, to which we will refer as $s_{\text{mec}}(\mathbf{Y}, \mathbf{X})$ and is obtained with [15]

$$s(\mathbf{Y}, \mathbf{X}) = \frac{1}{N_s N_h} \sum_{l=1}^{N_s} \sum_{k=1}^{N_h} \frac{\hat{P}_{kl}}{\hat{\sigma}_{kl}^2}, \quad (15)$$

where $\hat{P}_{kl} = \|\mathbf{X}_k^T \mathbf{s}_l\|^2$ estimates the signal power and $\hat{\sigma}_{kl}$ provides an estimate to the noise power at the k -th harmonic in the l^{th} spatially filtered signal. $\hat{\sigma}_{kl}$

is obtained by fitting a p -order autoregressive AR(p) model, with parameters $\{\hat{\alpha}_{l1}, \dots, \hat{\alpha}_{lp}, \hat{\sigma}_l^2\}$ to the data of each column (l) in the matrix $\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S} = \tilde{\Psi} \mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N \times N_s}$ [15], where $l \in \{1, \dots, N_s\}$. It can be computed as

$$\hat{\sigma}_{kl}^2 = \frac{\pi N \hat{\sigma}_l^2 / 4}{|1 + \sum_{m=1}^p \hat{\alpha}_{lm} \exp(-j2\pi m k f / F_s)|^2},$$

where $j = \sqrt{-1}$, F_s is the sampling frequency and f is the stimulation frequency.

The discrimination power of the statistic in (15) stems from its ability to incorporate the noise power estimate at the frequencies under consideration. So far, the score functions in CCA and MSI reflected the signal power only. Furthermore, we can write the noise covariance matrix as

$$\begin{aligned} \mathbf{C}_{\psi\psi} &= \mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \\ &= \mathbf{C}_{yy} (\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}) \\ &= \mathbf{C}_{yy} (\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1/2} \mathbf{T}^T \mathbf{T} \mathbf{C}_{yy}^{1/2}) \\ &= \mathbf{C}_{yy} (\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1/2} \mathbf{W}'_y \mathbf{P}^2 \mathbf{W}'_y{}^T \mathbf{C}_{yy}^{1/2}) \\ &= \mathbf{C}_{yy} (\mathbf{C}_{yy}^{-1/2} \mathbf{W}'_y (\mathbf{I}_{N_y} - \mathbf{P}^2) \mathbf{W}'_y{}^T \mathbf{C}_{yy}^{1/2}) \\ &= \mathbf{C}_{yy}^{1/2} \mathbf{W}'_y (\mathbf{I}_{N_y} - \mathbf{P}^2) \mathbf{W}'_y{}^T \mathbf{C}_{yy}^{1/2} \quad (16) \\ &= \mathbf{C}_{yy} \mathbf{W}_y (\mathbf{I}_{N_y} - \mathbf{P}^2) \mathbf{W}_y^T \mathbf{C}_{yy}. \quad (17) \end{aligned}$$

Multiplying both sides with $\mathbf{W}'_y{}^T$ from the left and \mathbf{W}_y from the right, yields the following relation

$$\mathbf{W}'_y{}^T \mathbf{C}_{\psi\psi} \mathbf{W}_y = \mathbf{I}_{N_y} - \mathbf{P}^2. \quad (18)$$

Therefore it is possible to diagonalize the noise covariance matrix with the canonical weights matrix \mathbf{W}_y , which is generally not orthogonal, as by definition $\mathbf{W}'_y{}^T \mathbf{C}_{yy} \mathbf{W}_y = \mathbf{I}_{N_x}$. Recall that the diagonalization of $\mathbf{C}_{\psi\psi} = \mathbf{Q}_\psi \mathbf{D}_\psi \mathbf{Q}_\psi^T$ in the original paper of the MEC method was done with eigendecomposition [15]. Should the EEG data be spatially white or pre-whitened, i.e. $\mathbf{C}_{yy} = \mathbf{I}_{N_y}$ before running the MEC procedure, and N_s is fixed to $2N_h$, then $\mathbf{W}_y = \mathbf{W}'_y = \mathbf{Q}_\psi$. This result provides another intuitive insight on the MEC filtering. When the original EEG signals is pre-whitened, MEC maximizes the canonical correlation coefficients while minimizing the noise energy since the smallest diagonal elements in \mathbf{D}_ψ and $(\mathbf{I}_{N_y} - \mathbf{P}^2)$ correspond to the largest canonical correlations.

3.4. Maximum Contrast Combination (MCC)

The goal of the maximum contrast combination (MCC) method is to find the linear spatial filter that maximizes the generalized Rayleigh quotient [15, 17]. Formally, we are after \mathbf{w} which maximizes $\lambda = \frac{\mathbf{w}^T \mathbf{C}_{yy} \mathbf{w}}{\mathbf{w}^T \mathbf{C}_{\psi\psi} \mathbf{w}}$ subject to $\|\mathbf{w}\|_2 = 1$. The true

covariance matrices are not known and thus they are substituted with their sample estimates. The sample noise covariance matrix $\mathbf{Q}_{\psi\psi}$ can be found the same way as in section 3.3. With the help of Lagrangian multipliers, one can show that λ attains its maximum with the dominant eigenvector of the matrix $\mathbf{C}_{\psi\psi}^{-1} \mathbf{C}_{yy}$, which can be rewritten with the result in (16) as

$$\mathbf{C}_{\psi\psi}^{-1} \mathbf{C}_{yy} = \mathbf{C}_{yy}^{-1/2} \mathbf{W}'_y (\mathbf{I}_{N_y} - \mathbf{P}^2)^{-1} \mathbf{W}'_y{}^T \mathbf{C}_{yy}^{1/2}. \quad (19)$$

Thus, $\lambda_i(\mathbf{C}_{\psi\psi}^{-1} \mathbf{C}_{yy}) = 1/(1 - \rho_i^2) = f_{\text{mcc}}(\rho_i)$, $\forall i \in \{1, \dots, \min(N_x, N_y)\}$, and consequentially, $\lambda_{\text{max}} = 1/(1 - \rho_{\text{max}}^2)$. The spatial filter \mathbf{w} which attains the maximum quotient can be found by normalizing the columns $\mathbf{C}_{yy}^{-1/2} \mathbf{W}'_y = \mathbf{W}_y$ with respect to the Euclidean norm, and picking the one corresponding to ρ_{max} . This proves that MCC and the CCA methods have exactly the same discrimination power since the function $f_{\text{mcc}}(\rho_1)$ is monotonically increasing in ρ_1 .

3.5. Canonical variates with autoregressive spectral estimation of noise (CVARS)

As stated earlier, the standard CCA method is able to reveal spatial coherence in high SNR regimes. When $\text{SNR} \ll 1$, however, the canonical coefficients mainly reflect the correlation between the noise and the assumed source signals, which often leads to erroneous detection as the separate contribution of the signal and the noise to the total values of ρ_i cannot be determined.

Therefore, we propose to estimate the noise at each frequency for each signal after spatial filtering with parametric spectral density estimation. That is, the noise power is estimated by fitting an autoregressive model to canonical variates after cleaning them from all energy at the SSVEP driving frequencies and their higher harmonics. The test statistic s_{cvars} is computed exactly as in (15) with the AR(p) models fitted on $\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}$, where $\mathbf{S} = \mathbf{Y} \mathbf{W}_y$ and $\mathbf{W}_y \in \mathbb{R}^{N_y \times N_x}$ is the canonical weighting matrix. Additionally, we observe that $\mathbf{X}^T \mathbf{S} = N \mathbf{W}_x \mathbf{P}$ and therefore $\hat{P}_{kl} = N \rho_l^2 (w_{kl,1}^2 + w_{kl,2}^2)$, where $w_{kl,1}$ and $w_{kl,2}$ denote the respective weight of the sine and cosine signals of the k^{th} harmonic in the l^{th} canonical variate.

3.6. Discussion

In the light of the previous analysis and findings, it is obvious that the scores of the CCA, MCC and MSI methods correspond to different functions of the canonical correlations (ρ_i), and their scores can all be considered special cases of s_{fcc} . Common to all of them is the spatial pre-whitening step of both, the EEG signals and the assumed SSVEP pure response, which is spatially white by construction. Though it is not

necessary to do so to obtain the scoring functions s_{cca} and s_{mcc} , these methods additionally involve applying the transform $\mathbf{C}_{yy}^{-1/2}$ to the left singular vectors of the coherence matrix in order to obtain the optimal spatial filter. The spatial filtering of the MEC and CVARS methods is accomplished with the eigenvectors of the noise covariance matrix and the canonical weighting matrix \mathbf{W}_y , respectively. Additionally, both involve spectral analysis of AR(p) models fitted on the spatially filtered data cleaned from the energy at the driving frequencies of the SSVEP. The rationale behind using (3.3) in [15, 19] is that the temporally colored noise in each column $\tilde{\mathbf{s}}_l$ in the matrix $\tilde{\mathbf{S}}$, can be modeled as discrete-time autoregressive random process of order p , such that $\sum_{k=0}^p \alpha_k \tilde{\mathbf{s}}[n-k] = u[n]$, where $u[n]$ is white Gaussian noise of power σ^2 and $\alpha_0 = 1$. As a result, this modeling allows to whiten the temporally colored noise, and produce an unbiased (or with small bias [19]) estimate for the noise power at each stimulation frequency and its higher harmonics. The CVARS method, therefore, involves whitening the data, both spatially and temporally, before it can provide the scoring function s_{cvars} . If the data is spatially pre-whitened before applying the MEC method, then results will be very similar to those of the CVARS method, but not exactly the same as N_s used to compute s_{mec} is governed by a fraction of the total energy in the noise signal and in case of CVARS, $N_s = 2N_h$ is fixed.

4. Material and methods

In order to compare the different detection methods, several experiments with volunteer subjects were conducted.

4.1. Subjects

A total of 10 healthy adults (1 female) aged 29.3 ± 5.5 (range 22 – 39) with normal or corrected-to-normal vision served as paid volunteer subjects in this study. During the experiments, the participants were seated 0.65 m away from an LCD monitor on a comfortable armchair in a slightly dimmed room. All participants gave their written informed consent. Participants were additionally asked to fill in pre- and post-questionnaires, that were meant to collect data about the level of tiredness before and after the experiment in addition to some demographical data.

Scalp EEG signals were recorded from 16 electrodes positioned according to the international extended 10/20 electrode system over the parieto-occipital scalp areas at P3, Pz, P4, PO9, PO7, PO3, POz, PO4, PO8, PO10, O9, O1, Oz, O2, O10 and Iz. Electrodes were referenced to the right earlobe and the ground electrode was positioned at FPz. The

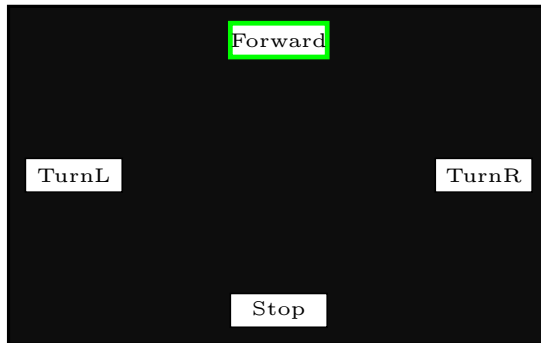


Figure 4: Stimulus presentation. In synchronization with the display refresh rate, four white rectangles were flickered on and off against a black screen at rates of 15 Hz (Forward), 12 Hz (TurnL), 10 Hz (TurnR) and 8.57 Hz (Stop). The labeling of the different rectangles serves robotic control applications. The size proportions between the flicker and the display are preserved in the figure.

signals were acquired with sampling rate of 512 Hz using g.USBamp acquisition system (g.tec medical engineering GmbH, Schiedlberg, Austria) and band-pass filtered at 0.5 – 60 Hz. The power line interference at 50 Hz was removed with a 4th order butterworth notch-filter with 48 – 52 Hz stop band. All electrodes were filled with highly conductive gel in order to reduce impedance.

This study is part of a larger project, which is approved by the Ethics Committee of the Faculty of Medicine of the Technical University of Munich (TUM).

4.2. Experimental paradigm

A 22" liquid-crystal display (LCD) monitor[‡] with 60 Hz refresh rate and 1280 × 1024 resolution was used to view the stimuli which consisted of four spatially distributed flickering rectangles presented simultaneously to participants. The driving

[‡] In addition to the LCD monitor, participants viewed stimuli via a head-mounted display (HMD) from Oculus VR, United States. The HMD has 60 Hz refresh rate as well. Each subject sequentially viewed the stimuli on both displays (i.e. the monitor and the HMD) with either their left eye only, right eye only, or both eyes. This resulted in a total of 2 × 3 different viewing conditions which were pseudo-randomized across subjects. Subjects were assigned either to finish all the HMD or the LCD conditions at first to minimize electrode displacement that might take place after mounting/unmounting the HMD. Two consecutive sessions per condition were recorded for later analysis. In total, each subject underwent 12 sessions, each of around 5 min duration. The whole experiment lasted around two hours including preparations and rest breaks between sessions. In order to provide results that can be compared with those in the literature, we will base our evaluation throughout this work solely on the two recording sessions obtained with binocular LCD viewing.

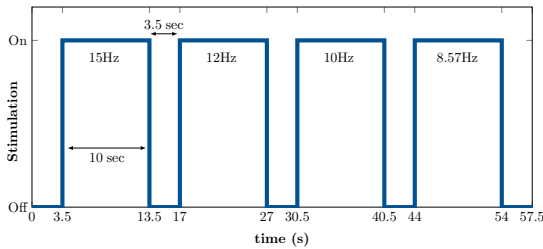


Figure 5: Trigger timing for one complete stimulation sequence. When stimulation is on, all flickering stimuli are presented concurrently. Each stimulus is highlighted for 10 seconds as shown with the stimulation train signal.

frequencies of the stimuli were chosen as integer divisors of the display refresh rate, namely 15, 12, 10 and 8.57 Hz. The chosen and *fixed* driving frequencies are known to evoke moderate to high SSVEP’s amplitude strength [29]. The spatial distribution of the stimuli is shown in figure 4. The EEG acquisition and visual stimulation were running on two different computers, and synchronized with the screen overlay control interface (SOCl) [30, 31].

Two recording sessions were performed, each started with a blank screen for around 15 seconds followed by the presentation of the flickering stimuli. During stimulation, participants were instructed to overtly sustain the spatial attention on the cued stimulus. Stimuli were highlighted in turn with a green rectangle as shown in figure 4 for 10 seconds followed by a rest period of 3.5 seconds, in which the screen went blank. The stimuli were cued in the descending order of their driving frequencies (i.e. the sequence will be top, left, right and bottom according to the stimuli constellation shown in figure 4). Detailed timing of one complete sequence is shown in figure 5. Each recording session consisted of five such full sequences.

5. Experimental Results

The accuracy for each of the different detection methods is highly influenced by the choice of the key system parameters, e.g N , T , N_h and F_s . In the following, we will firstly highlight the effect of each of these parameters individually on unsupervised CCA detection accuracy. We chose the CCA method here as the other scoring functions can be obtained from the canonical coefficients, canonical variates and the canonical weights, and thus it can serve as an indicator of the information gain/loss that accompanies parameter change. By fixing these parameters in the light of the empirical evaluation of CCA, the unsupervised detection accuracy of the CVARS is then compared to the state-of-the-art

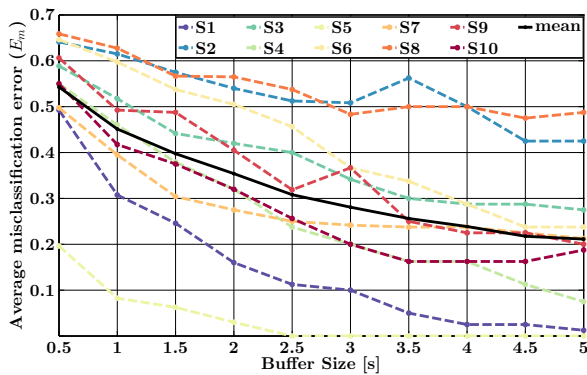


Figure 6: Average misclassification error for CCA computed as $E_m = 1 - \bar{P}_D$ for varying buffer length, $N_h = 1$.

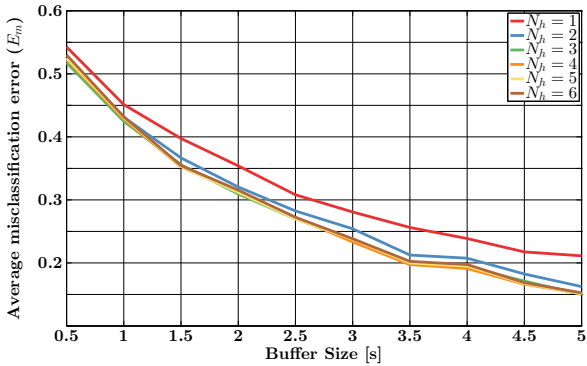
methods.

5.1. CCA results with varying key system parameters

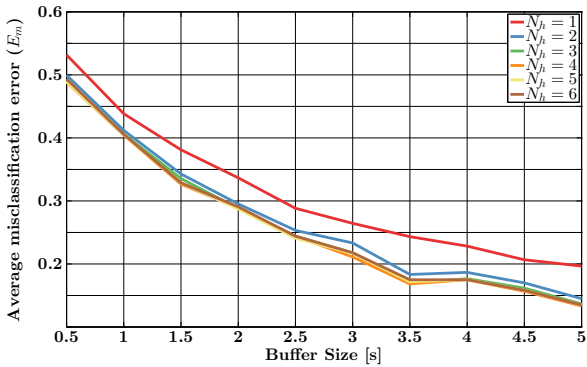
Labeled non-overlapping EEG data segments (i.e. $O = 0$) are extracted from the two available recording sessions per subject. The segment size was varied between 0.5 – 5 seconds with steps of 0.5 seconds, respectively yielding 200, 100, 60, 50, 40, 30, 20, 20, 20 and 20 segments, per stimulation frequency. Segments which were obtained during the idle state were not included in the evaluation since we consider unsupervised CCA at this stage.

Figure 6 shows the average misclassification error (E_m) for all subjects as a function of the buffer size, when the CCA method was used with $N_h = 1$ and $F_s = 512$ Hz. With larger buffer sizes, one can observe that misclassification errors for all subjects get suppressed due to enhanced SNR and more accurate estimates of \mathbf{Q}_{yy} and \mathbf{Q}_{xy} .

Figure 7(a) shows the average misclassification error of the CCA method (averaged over all subjects) for different number of harmonics N_h that ranges from 1 to 6 as a function of buffer length. For $N_h > 3$, no further improvement is observed, which cannot be explained by the fact that EEG data itself was lowpass-filtered with a cutoff frequency 60 Hz since the fourth harmonic of the maximum stimulation frequency is not rejected thereby. However, the fourth harmonic of the driving frequency $f_2 = 12$ Hz, lies within the stop band of the notch filter. In order to fully isolate the influence of the notch filter, we reevaluated the average misclassification error of the unsupervised CCA method (excluding f_2 from the analysis). Again, the results shown in figure 7(b), suggest that no further improvement for $N_h > 3$. The increased accuracy for $N_h > 1$ is however not statistically significant as it was also reported in [22]. The value of N_h is set to 3 throughout the remaining of this work.



(a)



(b)

Figure 7: Average misclassification error for CCA computed with different number of harmonics. For $N_h > 3$, no further improvement in accuracy is observed. Evaluation is based on (a) all stimulation frequencies (b) and excluding $f_2 = 12$ Hz.

As has been mentioned earlier, one can derive arbitrary scoring functions s_{fcca} from the canonical correlations ρ_i . Figure 8 shows the results for $s_{fcca} = \sum_1^k \rho_i$, while fixing $N_h = 3$, and $k \in \{1, 2, \dots, 2N_h\}$. The performance degradation of s_{fcca} for $k > 1$ suggests that the fluctuations in $\rho_i, \forall i > 1$ over time cannot be used in winner takes all (WTA) assignment as ρ_1 for SSVEP detection.

The effect of changing the sampling rate is shown in figure 9. Downsampling with a factor of 2 or 4, which respectively resembles sampling frequency of 256 and 128 Hz leads to accuracies that are comparable with the full data segments (with $F_s = 512$). Estimation of covariance matrices \mathbf{Q}_{yy} and \mathbf{Q}_{xy} is not, therefore, significantly affected by downsampling. This behavior suggests that adjacent samples in \mathbf{Y} are correlated (and they are) and that allowing for Δt between samples that is larger than the expected maximum correlation lag would not affect the obtained results. Going to downsampling factor of 8 ($F_s = 64$) deteriorates the performance significantly, as this would introduce

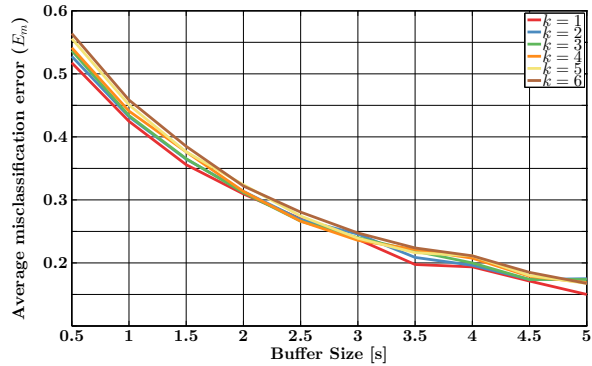


Figure 8: Average misclassification error for CCA averaged over all frequencies and subjects computed for different scoring function of ρ_i and $i \in \{1, \dots, N_x\}$.

aliasing and loss of information of the higher harmonics (with frequencies larger than $F_s/2$). Throughout the remaining of this work, we use $F_s = 512$ Hz. Sampling rates $F_s \in \{256, 128\}$ are expected to produce similar results in terms of the reported scores and the reported misclassification rates.

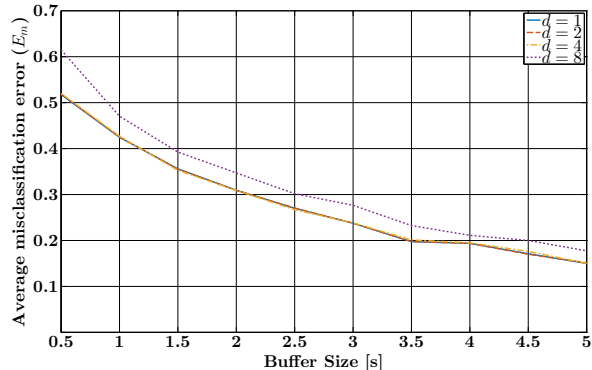


Figure 9: Average misclassification error for CCA averaged over all frequencies and subjects for different downsampling factors. Hardly any difference is noticed for $d \in \{1, 2, 4\}$. When the downsampling factor is 8 (the dotted purple line), significant reduction in accuracy is observed.

The estimates of the detection accuracies obtained so far using the non-overlapping data segments can be a bit misleading as classification is performed only on *homogeneous* EEG data segments, during which subjects attended to one single driving frequency. However, during online usage, EEG data collected in one segment can reflect two or more different states of user attention (e.g. user shifts his attention from one stimulus to another or from the idle state to one of the active states within the segment). This behavior becomes more probable with larger buffer sizes. To simulate this case, maximally overlapping

data segments (i.e. $T = 1$ sample) were continuously extracted from the data and used to plot the CCA score evolution of the different stimulation frequencies over time. The resulting segments thus contain both homogeneous and heterogeneous data. In order to additionally provide more insights about the inter-subject variability in figure 6, the score evolution will be shown for the subjects $S5$ and $S2$, whose CCA results were the best and the worst, respectively. CCA score evolution during the first full sequence (after viewing all stimulation frequencies once) for $(N, T) = (1024, 1)$ samples is shown in the upper row of figure 13. These plots show, to some extent, that during stimulation with f , the score $s(\mathbf{Y}, \mathbf{X}_f)$ increases over time and surpasses the scores of the other frequencies. Additionally, among the used stimulation frequencies, one can observe for each subject, that a specific frequency is somewhat dominant throughout the whole sequence (12 Hz for $S5$ and 10 Hz for $S2$), and to this specific frequency most of the faulty detections and false alarms can be attributed. While high scores for the dominant frequency (most likely due to interference from the alpha brain band waves, within which a peak can be observed in most subjects' EEG[32]) starts to appear during the idle state, they get suppressed (though not always) when subjects shift their visual attention to flickering light. Furthermore, one can categorize the SSVEPs of the two subjects into high and low SNR with respect to the obtained CCA score values, where $S2$ is the one with the low SNR.

5.2. Distribution of canonical correlations

In the following, we will refer to the canonical correlation values obtained when users attended to a specific stimulus with frequency f as the target canonical correlations (or target scores) for that frequency. Nontarget scores of a stimulation frequency f , on the other hand, refer to the values obtained when the user attended to other or no stimuli. By fixing $N_h = 3$ and $(N, T) = (1024, 1)$ samples, we computed the target and nontarget correlation coefficients for all stimulation frequencies in all recording sessions for all subjects. We assigned a data segment to a frequency f , if the most recent sample in the buffer was obtained when the corresponding stimulus was then cued for viewing.

From the histogram of all these values, we estimated the distribution of all target and nontarget canonical correlations which are shown in figure 10. These plots show that the difference between the distributions of the target and nontarget canonical correlations ρ_i is most pronounced for ρ_1 . Most importantly, we could see that the typical means (for all canonical correlations) per stimulation frequency differ significantly, in a way that reflects the general

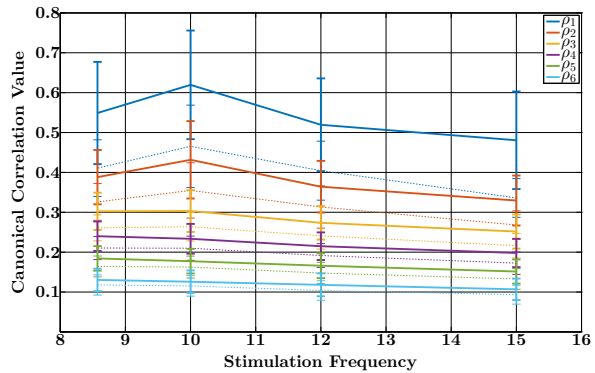


Figure 11: Mean and standard deviation of target (solid) and nontarget (dotted) ρ_i computed for the different stimulation frequencies.

power density of EEG data which, similar to pink noise, exhibits a characteristic $1/f$ profile [32, 33]. The mean and standard deviation of all target and nontarget canonical correlations are shown in figure 11 as a function of stimulation frequency f . Besides the $1/f$ profile, we can observe a peak at 10 Hz, which stems from interference of the dominant alpha brain waves at the same frequency. The distributions of the target and nontarget ρ_i 's clearly justify why s_{fcca} defined as the sum of canonical correlations did deteriorate detection performance when compared to s_{cca} , as the bias towards the low frequencies and the subject-dependent peaks in the alpha band, increases by adding further correlations to the value of ρ_1 . Therefore, the CCA scores need to be scaled differently for each frequency in order to correct for the observed bias. This scaling is done efficiently in the CVARS method.

5.3. Comparison of the different methods

The same non-overlapping EEG segments from section 5.1 were used to compare the CCA, MCC, MSI, MEC and CVARS methods. The mean misclassification error averaged over all subjects is shown in figure 12 for $N_h = 3$ and $F_s = 512$ Hz. MEC and CVARS were used with AR(7) model. By visual inspection, the results of the CCA, MCC and MSI methods don't differ significantly. The MEC and CVARS methods outperform CCA for all buffer lengths, except for 0.5 s buffers, in which case, all methods have comparable accuracies. The CVARS method performs slightly better than the MEC for almost all buffer lengths.

Furthermore, the upper three rows in figure 13 compare the CCA, MEC and CVARS with respect to their score evolution during the first full stimulation sequence with $(N, T) = (1024, 1)$ samples. CVARS and MEC methods show significant improvement over CCA for the subject with the low SNR ($S2$), especially

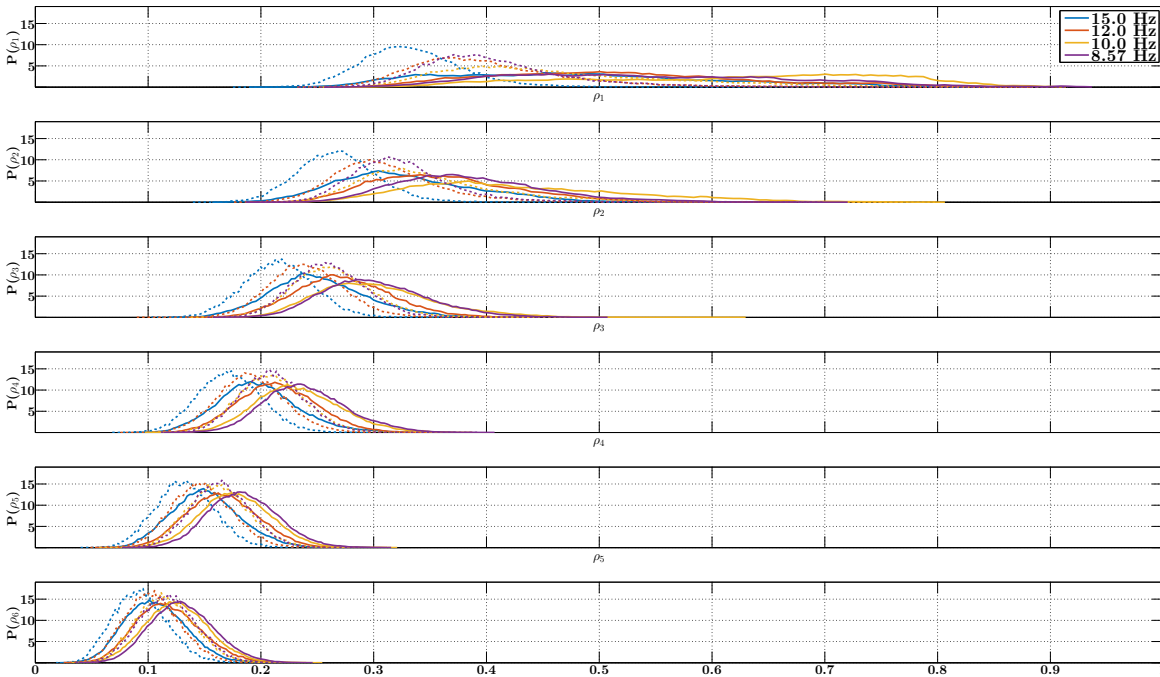


Figure 10: Distribution of target (solid) and nontarget (dotted) canonical correlations averaged over all subjects from overlapping EEG segments. For instance, the distributions of target ρ_i for $f_1 = 15$ Hz were obtained by computing the canonical correlations of EEG segments $\mathbf{Y} \in \mathbb{R}^{1024 \times 16}$ where $g(\mathbf{Y}) = f_1$ with $s_{cca}(\mathbf{Y}, \mathbf{X}_{f_1})$. The nontarget distributions were obtained from segments with $g(\mathbf{Y}) \neq f_1$.

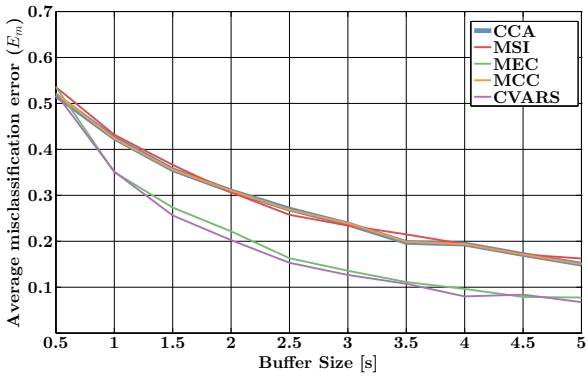


Figure 12: Average misclassification accuracy averaged over all frequencies and subjects. As expected, classification accuracies for CCA (plotted with thick line to highlight this fact) and MCC are the same.

for the stimulation frequency $f_1 = 15$ Hz and less for $f_2 = 12$ Hz. Less or no improvement can be observed for the subject with the high SNR ($S5$). From the score evolution ($N = 1024, T = 1$) data, we have computed the average accuracy per subject over all stimulation frequencies (excluding the idle state) for the unsupervised CCA, standard MEC and CVARS. A repeated measures ANOVA determined

that mean classification accuracy differed statistically significantly between the three methods ($F(2, 18) = 19.902, p < 0.0001, \text{partial } \eta^2 = 0.689$). Post hoc tests using the Bonferroni correction revealed statistically significant ($p = 0.005$) improvement of unsupervised CVARS (mean \pm std: 0.739 ± 0.099) over unsupervised CCA (mean \pm std: 0.64 ± 0.137). Furthermore, we found a trend for unsupervised CVARS being superior over unsupervised standard MEC, which given the current number of subjects though is not significant ($p = 0.069$). Unsupervised standard MEC (mean \pm std: 0.72 ± 0.107) performed significantly better than the unsupervised CCA ($p = 0.003$).

The effect of choosing longer buffers ($N = 2560$) with CVARS method is shown in the lower panel of figure 13. We can easily observe that fluctuations in the score function are reduced for both $S2$ and $S5$, with the side effect of introducing delays into the system (rise-up and decaying delays). This, however, did not guarantee reliable SSVEP detection for $S2$.

5.4. Supervised CVARS method for SSVEP detection

The unsupervised detection methods discussed so far provided estimates of the driving frequency in a WTA fashion. This way, the system is always in the control loop and the idle state is never reached. Furthermore, the target and nontarget scores distributions in

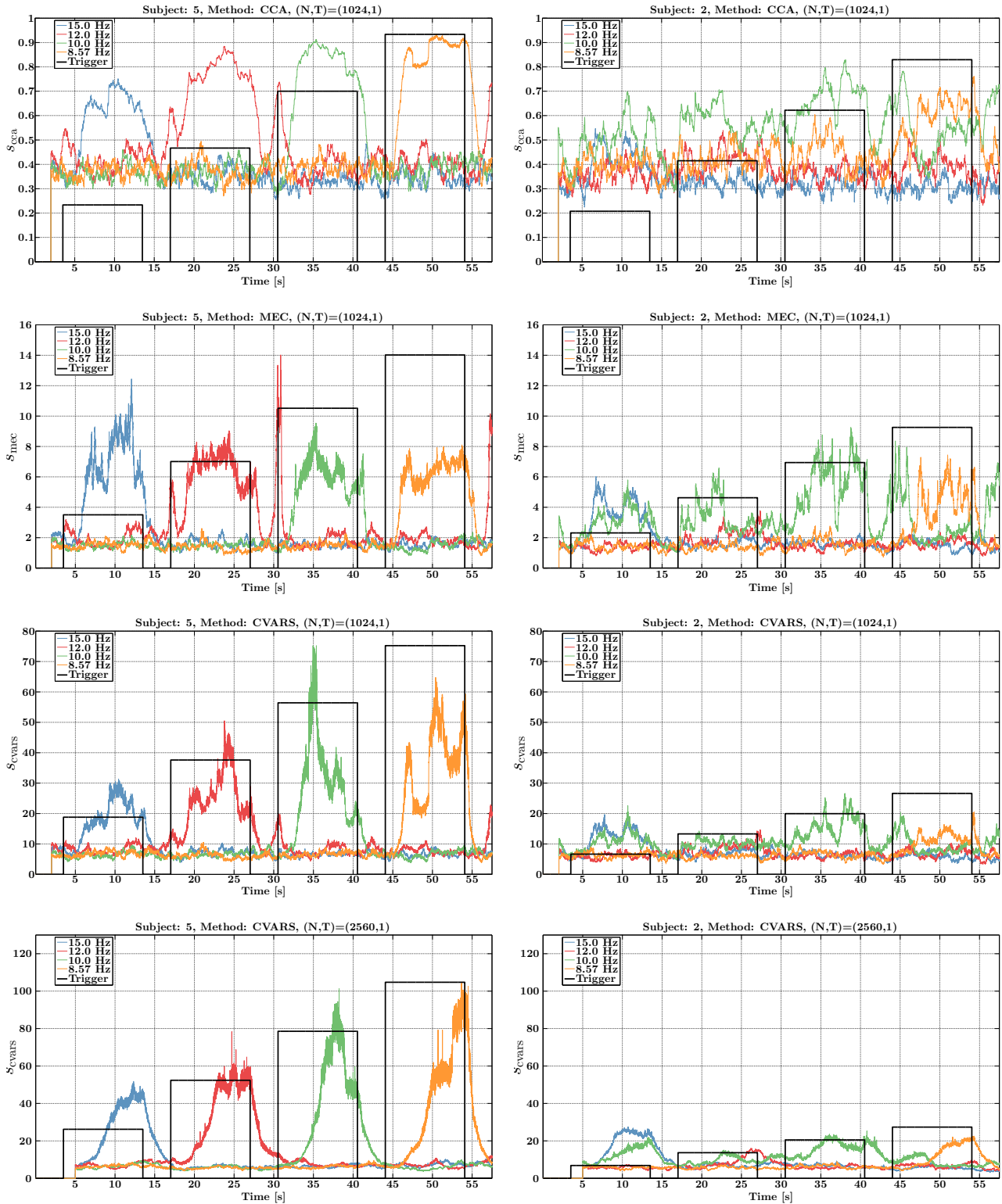


Figure 13: Score evolution for S_5 (left) and S_2 (right) obtained during the first stimulation sequence of the first recording session for buffers of 2 seconds (except for the last row this was 5 s), $T = 1$ sample and $N_h = 3$ with the CCA, MEC, and CVARS methods. AR(7) models were used when necessary. The trigger is shown as a staircase with 4 levels, each level corresponds to one stimulation frequency. For example, level 2 corresponds to $f_2 = 12$ Hz. The zero level corresponds to the idle state. The fourth level (i.e. for f_4) is scaled to the largest score obtained for better visibility.

figures 10 and 13 suggest that the simple argmax function in (5) will provide faulty detections due to the large visible difference in the mean scores for the different driving frequencies, and the resulting bias towards the lower frequencies. Learning a simple threshold from labeled data (obtained from a training session for each user), although can reduce the rate of false alarms and faulty detections, can lead also to a larger rate of misses as can easily be seen for CCA from the plots in figure 10 and for all methods in figure 13. For instance setting a CCA threshold of 0.5 in figure 10 would result in a high probability of miss, as can be computed from the area under the curves of the target distributions for $\rho_1 < 0.5$. It can be additionally observed that it is more probable to miss the detection of higher frequencies than the lower ones. Linear Discriminant Analysis (LDA) lends itself naturally to such a problem, where labeled score vectors are obtained from one training session and an LDA classifier is thereby learned. In online sessions, the scores are computed as usual to produce the vector $\mathbf{s} \in \mathbb{R}^{N_f \times 1}$. The LDA classifier is then applied on \mathbf{s} to provide the final estimate of the hidden attended stimulus. We summarize the comparison between the unsupervised and supervised CVARS with $(N, T) = (1024, 128)$ in the form of the confusion matrices shown in figure 14, for subjects S5 and S2, in addition to the average over all subjects. We chose CVARS here as it provided the best results in the unsupervised case. The confusion matrices were computed with each method applied on the two available recording sessions. In case of the supervised CVARS, a classifier was learned from each session and applied on the other one, and results therefrom were averaged (i.e. a total of two classifiers were used per subject). These results clearly show a reduction in the rate of false alarms on the expense of higher probability of misses. The probability of miss and that of correct detections are quite uniform with respect to the stimulation frequencies, on the contrary to what is expected if a single threshold were used. Furthermore, a large portion of the probability of miss in the supervised CVARS replaces some of the wrong detections in the unsupervised case. Falsely detecting one control state as the idle state is generally favored over confusing it with another control state.

5.5. Discussion

From the results in the previous subsections, we can see that there are conflicting factors that affect the detection accuracy of SSVEP. On one hand, accuracy is a monotonically increasing function of the buffer size (i.e. N), should the buffer contain only homogeneous data, which is not the most probable case in online applications. On the other hand, larger buffer sizes introduce delays, as the old samples which contain no

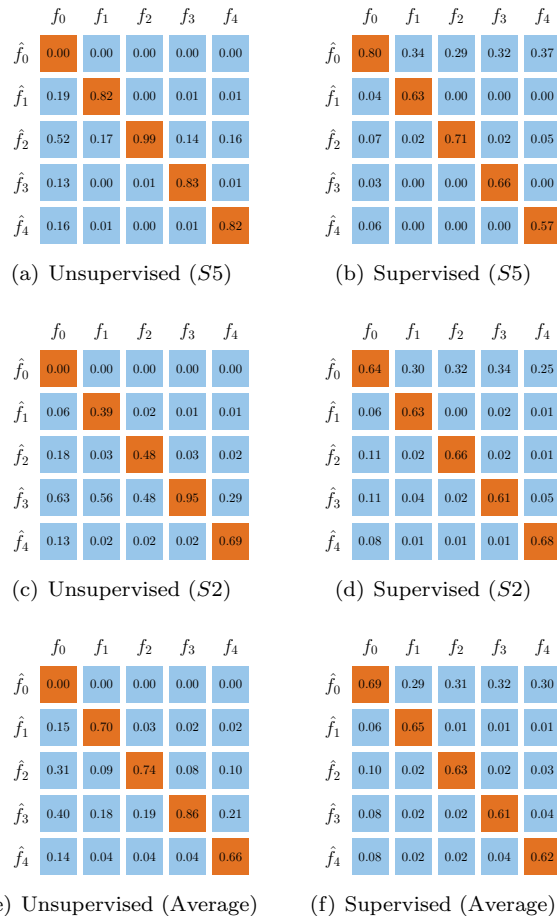


Figure 14: Confusion matrices for the unsupervised and supervised CVARS methods. The supervised CVARS method produces reliable results for all participants including S2, whose data has been shown to have low SNR. This proves the ability of supervised CVARS to deal with wide range of SNR levels.

information about the currently attended stimulus still inhabit the buffer \mathbf{Y} and contribute to the different values $s(\mathbf{Y}, \mathbf{X}_f)$, leading to false alarms and faulty detections.

Comparing figures 6 and 15, which respectively show the misclassification results of non-overlapping segments as a function of buffer length for CCA ($N_h = 1$) and CVARS ($N_h = 3$), we can see that buffer lengths, at which acceptable accuracies are obtained, differ from one subject to another, regardless of the detection method used. Therefore, a trade-off between accuracy and speed should be optimized for each subject, based on a short training session. The same session is also used to learn the LDA classifier from the obtained scores.

Additionally, with our theoretical analysis and empirical results, we have shown that CCA and MCC

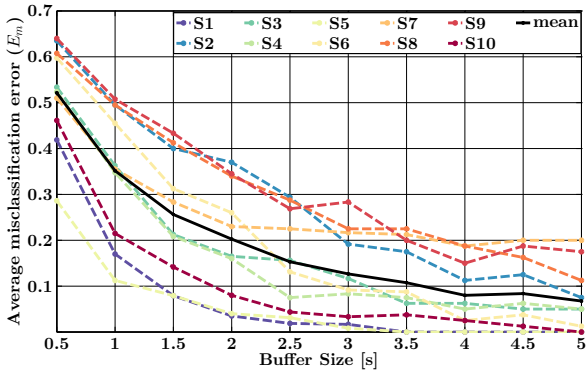


Figure 15: Average CVARS-method misclassification error computed as a function of buffer length for non-overlapping segments.

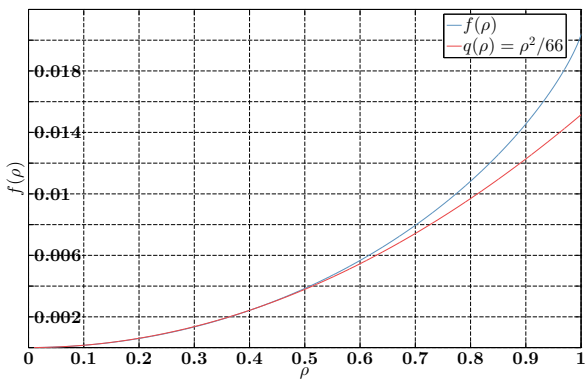


Figure 16: The function $f(\rho)$ qualitatively shows the relatively large contribution of the maximum canonical correlation in the final MSI score, especially when $\rho_1 > 0.5$. For $\rho_1 \leq 0.5$, $s_{\text{msi}} \approx \sum_{i=1}^{2N_h} \rho_i^2$.

give exactly the same estimates and thus the same misclassification error rates. MSI and CCA have shown similar misclassification error rates, suggesting that ρ_1 plays a major role in calculating s_{msi} . This can be seen by rewriting (13) as $s_{\text{msi}} = \sum_{i=1}^{2N_h} f(\rho_i)$, where

$$f(\rho) = \frac{(1 + \rho) \log(1 + \rho) + (1 - \rho) \log(1 - \rho)}{P \log(P)}. \quad (20)$$

Figure 16 plots $f(\rho)$ in the range of $[0, 1]$ in addition to the function $q(\rho) = \rho^2/C$, where C is a constant. The quadratic function $q(\rho)$ approximates $f(\rho)$ relatively well when $\rho < 0.5$. Therefore, for the case when all $\rho_i < 0.5$, the MSI score function (and its quadratic approximation) gives more weight to greater ρ_i 's. On the other hand, when $\rho_1 > 0.5$, the contribution of ρ_1 becomes more emphasized as the difference $f(\rho_1) - q(\rho_1)$ grows very rapidly. The latter case is the more probable case given the empirical values of target ρ_i 's shown in figure 10.

Furthermore, we have shown that MEC outperforms CCA. This result differs from what was shown

in [34]. However, the score function s_{mec} used there was defined with $\hat{\sigma}_{kl} = 1$, which ignores the noise power at the stimulation frequencies and consequently MEC scores will be biased towards the low frequencies. This reaffirms the need to scale the different correlations with regard to estimates about the noise power, which is done efficiently with test statistics used in standard MEC and CVARS procedures. In [15], a test statistic similar to s_{mec} is used with the MCC spatial filter, where the spatial filter is obtained from a subset of the columns in \mathbf{W}_y . Again, this statistic should give similar results to the CVARS when the data is pre-whitened before applying the MCC filter.

Alternatively, Yin et al. [35] proposed the supervised CCA-RV method for the same purpose of reducing the variability in the final target scores of the different stimulation frequencies where $s_{\text{cca-rv}}$ was computed with

$$s_{\text{cca-rv}}(\mathbf{Y}, \mathbf{X}_f) = \frac{s_{\text{cca}}(\mathbf{Y}, \mathbf{X}_f) - s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)}{s_{\text{cca}}(\mathbf{Y}, \mathbf{X}_f) + s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)}, \quad (21)$$

where $s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)$ is the mean nontarget scores of a frequency f computed from a training session. Since the CCA-RV method was published very recently, we could not fully and fairly compare it to the CVARS method, especially also that real-time biofeedback mechanisms were employed in [35]. However, we claim that the supervised CVARS method described in this paper should outperform the supervised CCA-RV (when we ignore the effect of the biofeedback). Firstly, CVARS has been shown in the current paper to outperform CCA when results for all subjects were averaged and therefore plugging the CVARS scores instead of CCA in (21) is expected to provide, on average, better results. Additionally, LDA learns from the available training session the optimal mapping from the CVARS scores to the stimulation frequencies as it takes into account not only the mean values as it is the case in [35] but also the variances and covariations of the individual scores.

Throughout this work, we had a number of channels $N_y = 16$, which was larger than the number of signals assumed in the source model $N_x = 2N_h = 6$. Reversing this relation does not affect the obtained results. For the CCA method, the number of canonical correlations is upper bounded by $\min(N_x, N_y)$, which will be N_y in this case. The eigenvalues in the MSI method, will be defined the same way as in section 3.2 except that there are $N_x - N_y$ eigenvalues which have the value 1. The CVARS method, in this case, will provide no dimension reduction, as the number of canonical variates will be the same as the original EEG channels. On the contrary, the MEC method can always produce less output channels than available in the original EEG signals. Friman et al. [15] reported typical values for $N_y = 6$ and $N_s \in \{4, 5\}$ which are

comparable to $2N_h = 4$. For our dataset, we got typical values of $N_s \in \{14, 15, 16\}$ for $N_y = 16$ and $N_h = 3$. The result we proved here regarding the conditions when $\mathbf{W}_{\text{MEC}} = \mathbf{W}'_y$ makes the CVARS method more consistent with the assumptions about the the number of source model signals.

6. Conclusion

Detection of SSVEP in continuous EEG signals lies under the general problem of detecting sinusoids in noisy measurements, a problem that has been thoroughly investigated in the array signal processing field. We have theoretically shown the conditions in which state-of-the-art SSVEP detection methods share similar spatial filters, a step required to enhance the overall signal-to-noise ratio. The equivalence of the discrimination power of the MCC and CCA methods has been proven and it was conjectured that MSI should have very similar results as well. Empirical evaluations supported these results. The methods CCA, MCC and MSI rely on a single metric that is computed from the canonical correlations ρ_i to provide an estimate about the stimulation frequency, to which a user is attending. Thereby they fail to provide reliable estimates when the signal is lost in the noise floor. The MEC and the hereby proposed CVARS methods, in addition to the step of spatial filtering, base their discrimination upon the estimated signal and noise powers at each considered frequency (the fundamental stimulation frequency and its higher harmonics). We have shown that the CVARS and the MEC scores are the same, given that the EEG signal is spatially whitened before running MEC algorithm and N_s is rather fixed to $\min(N_y, 2N_h)$, i.e. number of canonical correlations. The CVARS method slightly outperformed the standard MEC method, which typically requires no spatial pre-whitening as reported in [15]. Finally, we have shown that the supervised CVARS method based on a short training session can be used to learn a mapping function rather than the maximizer (argmax) that estimates the hidden driving frequency and the idle state from the obtained scores, reliably and accurately. The training session should also serve the purpose of finding the optimal buffer size for a specific subject to be used in online applications.

Acknowledgments

This work is supported in part by the VERE project within the 7th Framework Programme of the European Union, FET-Human Computer Confluence Initiative, contract number ICT-2010-257695 and the Institute for Advanced Study, Technical University of Munich

(TUM).

Appendix A. Relation between MSI and CCA

Finding the eigenvalues of the matrix \mathbf{R} defined in (12) involves solving the characteristic equation $\det(\mathbf{R} - \lambda\mathbf{I}) = 0$, and thus

$$\det \left(\begin{bmatrix} \mathbf{I}_{N_x} - \lambda\mathbf{I}_{N_x} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{I}_{N_y} - \lambda\mathbf{I}_{N_y} \end{bmatrix} \right) = 0$$

Without loss of generality, we assume in the following that $N_y \geq N_x$. Using the Schur determinant identity we can write

$$\begin{aligned} 0 &= \det(\mathbf{I}_{N_y} - \lambda\mathbf{I}_{N_y}) \cdot \det \left((1 - \lambda)\mathbf{I}_{N_x} - \frac{1}{1 - \lambda}\mathbf{T}\mathbf{T}^T \right) \\ &= (1 - \lambda)^{N_y} \cdot \det \left(\frac{1}{1 - \lambda} \left((1 - \lambda)^2 \mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T \right) \right) \\ &= (1 - \lambda)^{N_y} \cdot \frac{1}{(1 - \lambda)^{N_x}} \cdot \det \left((1 - \lambda)^2 \mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T \right) \\ &= (1 - \lambda)^{N_y - N_x} \cdot \det \left((1 - \lambda)^2 \mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T \right), \end{aligned}$$

which means that either $\lambda = 1$ or $(1 - \lambda)^2$ is one of the eigenvalues of $\mathbf{T}\mathbf{T}^T$. Consequently, there are exactly $(N_y - N_x)$ eigenvalues of \mathbf{R} that take the value 1. The remaining $2N_x$ eigenvalues can be related to the canonical correlations (subsection 3.1) with $\lambda = 1 \mp \rho_i$, where $i = 1, \dots, N_x$.

References

- [1] Ken Nakayama and Manfred Mackeben. Steady state visual evoked potentials in the alert primate. *Vision Research*, 22(10):1261–1271, 1982.
- [2] Christoph S Herrmann. Human EEG responses to 1-100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, 137(3-4):346–353, apr 2001.
- [3] Athanasios P Liavas, George V Moustakides, Günter Henning, Emmanuil Z. Psarakis, and Peter Husar. A periodogram-based method for the detection of steady-state visually evoked potentials. *IEEE Transactions on Biomedical Engineering*, 45(2):242–248, 1998.
- [4] Almudena Capilla, Paula Pazo-Alvarez, Alvaro Darriba, Pablo Campo, and Joachim Gross. Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS one*, 6(1):e14543, 2011.
- [5] S. T. Morgan, J. C. Hansen, and S. A. Hillyard. Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10):4770–4774, may 1996.
- [6] Yu Te Wang, Yijun Wang, Chung Kuan Cheng, and Tzyy Ping Jung. Developing stimulus presentation on mobile devices for a truly portable SSVEP-based BCI. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 5271–5274, 2013.

- [7] Chuan Jia, Honglai Xu, Bo Hong, Xiaorong Gao, Zhiguang Zhang, and Shangkai Gao. A human computer interface using SSVEP-based BCI technology. *Foundations of Augmented Cognition*, 4565:113–119, 2007.
- [8] Gido Hakvoort, Boris Reuderink, and Michel Obbink. Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. Technical Report Technical Report TR-CTIT-11-03, Centre for Telematics and Information Technology University of Twente, Enschede, 2011.
- [9] Ming Cheng, Xiaorong Gao, Shangkai Gao, and Dingfeng Xu. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering*, 49(10):1181–1186, 2002.
- [10] Kian B Ng, Ross Cunningham, and Andrew P Bradley. Enhancing the classification accuracy of steady-state visual evoked potential-based Brain-Computer interface using component synchrony measure. *IEEE world congress on computational intelligence*, pages 1–6, 2012.
- [11] J D Victor and J Mast. A new statistic for steady-state evoked potentials. *Electroencephalography and clinical neurophysiology*, 78(5):378–388, 1991.
- [12] Matthias M. Müller, Terence W. Picton, Pedro Valdes-Sosa, Jorge Riera, Wolfgang A. Teder-Sälejärvi, and Steven A. Hillyard. Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Cognitive brain research*, 6(4):249–261, apr 1998.
- [13] Simon P. Kelly, Edmund C. Lalor, Ciarán Finucane, Gary McDarby, and Richard B Reilly. Visual spatial attention control in an independent brain-computer interface. *IEEE transactions on biomedical engineering*, 52(9):1588–1596, sep 2005.
- [14] Thomas Meigen and Michael Bach. On the statistical significance of electrophysiological steady-state responses. *Documenta Ophthalmologica*, 98(3):207–232, 2000.
- [15] Ola Friman, Ivan Volosyak, and Axel Gräser. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE transactions on biomedical engineering*, 54(4):742–50, apr 2007.
- [16] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. *IEEE Transactions on Biomedical Engineering*, 53(12):2610–2614, jun 2006.
- [17] Niya Wang, Tianyi Qian, Qing Zhuo, and Xiaorong Gao. Discrimination between idle and work states in BCI based on SSVEP. In *Proceedings of the 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, volume 4, pages 355–358, 2010.
- [18] Yangsong Zhang, Peng Xu, Kaiwen Cheng, and Dezhong Yao. Multivariate synchronization index for frequency recognition of SSVEP-based brain-computer interface. *Journal of neuroscience methods*, 221:32–40, jan 2014.
- [19] C E Davila, R Srebro, and I a Ghaleb. Optimal detection of visual evoked potentials. *IEEE transactions on biomedical engineering*, 45(6):800–803, 1998.
- [20] D Regan. Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and clinical neurophysiology*, 20(3):238–248, 1966.
- [21] Fahrettin F. Gonen and Gleb V. Tcheslavski. Techniques to assess stationarity and gaussianity of EEG: An overview. *International Journal of Bioautomation*, 16(2):135–142, 2012.
- [22] Guangyu Bin, Xiaorong Gao, Zheng Yan, Bo Hong, and Shangkai Gao. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *Journal of neural engineering*, 6(4):046002, aug 2009.
- [23] Hongya Ge, Ivars P. Kirsteins, and Xiaoli Wang. Does canonical correlation analysis provide reliable information on data correlation in array processing? In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2113–2116, 2009.
- [24] Harold Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3-4):321–377, 1936.
- [25] David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, dec 2004.
- [26] Louis L. Scharf and John K. Thomas. Wiener filters in canonical coordinates for transform coding, filtering, and quantizing. *IEEE Transactions on Signal Processing*, 46(3):647–654, mar 1998.
- [27] Ake Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–579, 1973.
- [28] Cristian Carmeli. *Assessing cooperative behavior in dynamical networks with applications to brain data*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2006.
- [29] Yangsong Zhang, Peng Xu, Tiejun Liu, Jun Hu, Rui Zhang, and Dezhong Yao. Multiple frequencies sequential coding for SSVEP-based brain-computer interface. *PLoS one*, 7(3):e29519, 2012.
- [30] Christoph Kapeller, Kyousuke Kamada, Hiroshi Ogawa, Robert Prueckl, Josef Scharinger, and Christoph Guger. An electrocorticographic BCI using code-based VEP for control in video applications: a single-subject study. *Frontiers in systems neuroscience*, 8, 2014.
- [31] C. Kapeller, C. Hintermüller, M. Abu-Alqumsan, R. Prueckl, A. Peer, and C. Guger. A BCI using VEP for continuous control of a mobile robot. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 5254–5257, 2013.
- [32] Stefan Haufe. *Towards EEG source connectivity analysis*. PhD thesis, Technische Universität Berlin, 2012.
- [33] Charmaine Demanuele, Christopher J James, and Edmund Js Sonuga-Barke. Distinguishing low frequency oscillations within the 1/f spectral behaviour of electromagnetic brain signals. *Behavioral and brain functions : BBF*, 3:62, 2007.
- [34] Wenya Nan, Chi Man Wong, Boyu Wang, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai. A comparison of minimum energy combination and canonical correlation analysis for SSVEP detection. In *2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011*, number 3, pages 469–472, Cancun, apr 2011.
- [35] Erwei Yin, Zongtan Zhou, Jun Jiang, Yang Yu, and Dewen Hu. A Dynamically Optimized SSVEP Brain-Computer Interface (BCI) Speller. *IEEE Transactions on Biomedical Engineering*, 62(6):1447–1456, 2015.