# Maximizing Information Transfer in SSVEP-Based Brain–Computer Interfaces

Malte Sengelmann*, Andreas K. Engel, and Alexander Maye

***Abstract*—Compared to the different brain signals used in brain–computer interface (BCI) paradigms, the steady-state visually evoked potential (SSVEP) features a high signal to noise ratio, enabling reliable and fast classification of neural activity patterns without extensive training requirements. In this paper, we present methods to further increase the information transfer rates (ITRs) of SSVEP-based BCIs. Starting with stimulus parameter optimizations methods, we develop an improved approach for the use of Canonical correlation analysis and analyze properties of the SSVEP when the user fixates a target and during transitions between targets. These transitions show a negative effect on the system's ITR which we trace back to delays and dead times of the SSVEP. Using two classifier types adapted to continuous and transient SSVEPs and two control modes (fast feedback and fast input), we present a simulated online BCI implementation which addresses the challenges introduced by transient SSVEPs. The resulting system reaches an average ITR of 181 Bits/min and peak ITR values of up to 295 Bits/min for individual users.**

***Index Terms*—Canonical correlation analysis with partially fixed spatial filter (CCAFSF), diagonal quadratic discriminant analysis (DQDA), standardized mean of a contrast variable (SMCV).**

## I. Introduction

STEADY-STATE visually evoked potentials (SSVEPs) represent one of the most investigated and robust paradigms for brain–computer interfaces (BCIs). It exploits the effect that directing overt or covert [1] attention to flickering visual stimuli enhances electrical potentials over the visual cortex evoked at the flicker frequency. Tagging visual stimuli by different frequencies, phases, or temporal patterns of visual flicker and classifying the respective features in the recorded brain signal thus allows to decode to which of the stimuli the attention of the user is directed. Depending on the coding dimension, SSVEP BCIs can be categorized as time-division multiple access (TDMA), frequency- (FDMA), code- (CDMA), or space-division multiple access (SDMA) SSVEP BCI [2]. The usability, reliability, and speed of an SSVEP BCI are mainly determined by the parameters of the stimulation as well as the properties of the classification method. Stimulus parameters comprise color [3], contrast [4], duty cycle [5], movement [6], [7], stimulation frequencies [8], as well as different wave forms [9], [10]. Classification accuracy has been increased using different filter and feature extraction techniques like fast Fourier transform (FFT) [11], common spatial pattern [12], similarity of background [13], canonical correlation analysis (CCA) [14], filter bank CCA (FBCCA) [15], multichannel CCA [16], or joint frequency-phase modulation (JFPM) [17].

The currently fastest SSVEP BCIs achieve average information transfer rates (ITRs) between 108 and 267 Bits/min ([9], [15]–[19]). Most recent improvements to SSVEP detection ([15]–[17]) center around optimizations of the standard CCA approach [14]. Chen *et al.* [15] (151.18 Bits/min) improved the reference signal used in [14] by extracting the independent information embedded in the harmonics of the SSVEP more efficiently via FBCCA. Nakanishi *et al.* [16] (166.91 Bits/min) trained the reference signals from electroencephalogram (EEG) data and improved the efficiency of the spatial filter component of the CCA by simultaneously projecting onto multiple spatial filters calculated from the trained reference, test, and ideal sinusoid signals. Recently, Chen *et al.* [17] adopted parts of the analysis in [16] and added JFPM stimulus encoding which considerably improved the ITRs achievable by modern SSVEP BCI (from 166.91 Bits/min [16] to 267 Bits/min). The JFPM stimulus design aims at efficiently decoding a large number of frequency targets in a narrow frequency band (steps of 0.2Hz in [17]). In JFPM, the contrast between targets is maximized by equally spaced phases between adjacent frequencies.

The aim of our study was to explore the limits of the ITR of FDMA SSVEP BCIs and to approach this limit by a combined optimization of stimulation and classification methods. To this end, we split the system's signal processing pathway into its main components and analyzed those separately using idealized and simplified test setups. This investigation revealed their individual properties and performance optima.

On the methodological side, our optimization approach comprised the construction of a real time capable, high-precision stimulation hardware, an improved application of the CCA, and a quantification of the ITR gain of a new fast input (FI) mode in comparison to the conventional fast feedback (FF) mode of existing FDMA SSVEP BCIs.

These improvements were complemented at the subject level by optimizing the stimulation frequencies and feature

Manuscript received August 19, 2015; accepted April 17, 2016. Date of publication April 27, 2016 This work was supported by grants from the European Union (FP7-ICT-270212, ERC-2010-AdG-269716) and the Deutsche Forschungsgemeinschaft (GRK 1247/1/2). *Asterisk indicates in corresponding author.*

*M. Sengelmann is with theDepartment of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany (e-mail: m.sengelmann@uke.de).

A. K. Engel and A. Maye are with the Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf.

Digital Object Identifier 10.1109/TBME.2016.2559527

extraction methods for each user, determining the corresponding upper limit of the ITR, analyzing the different properties of the continuous and transient SSVEP, and deriving corresponding signal filters.

Recent SSVEP BCI implementations ([9], [15], [16], [18], [19]) usually operate in closed loop (proposed by Yuan *et al.* [20]) FF mode and do not consider the hard- and software delays of the setup. We show that ITR measurements based on an FF mode are biased by those delays and thus not directly comparable without quantification of the delays.

Combining the results led to a flexible BCI setup which features ITR optimization for both modes of operation (FI and FF) and achieves a mean ITR of 181 Bits/min and a peak ITR of 295 Bits/min measured at the output of the Level 1 BCI control module [24] in a simulated online experiment without feedback.

## II. Maximizing the ITR

The ITR is a measure for the amount of information transferred per time over an unreliable transmission line. It reflects a tradeoff between a cost resulting from the time required to transmit a single character and a gain represented by the amount of information a character holds and the probability to correctly transmit and receive it. Shannon and Weaver [21] expressed the ITR by (1), where B is the bit rate in bits per character, $N$ is the number of characters, TP is the classification accuracy, and $\Delta t_{\mathrm{ITR}}$ is the time required to transmit each character

$$\begin{aligned} B = {} & ld\,(N) + \mathrm{TP}\; ld(\mathrm{TP}) \\ & + (1-\mathrm{TP})\; ld\,(\frac{1-\mathrm{TP}}{N-1}) \\ \mathrm{ITR} = {} & B^{*}(60/\Delta t_{\mathrm{ITR}}). \end{aligned} \tag{1}$$

The three general options to increase the ITR are to decrease the time required to transmit each character, to increase the information each character holds, and to increase the recognition accuracy of each received character. However, these parameters, which need to be optimized in order to achieve maximum performance, are not independent.

### A. Optimizing the Stimulation Parameters

From (1), it follows that the ITR grows with the number of characters or target classes $N$. The logarithmic relation suggests that it is not necessary to maximize $N$ at all costs, but that still a reasonably large number of classes should be employed. A relatively simple way of increasing the number of targets in FDMA BCIs is to use the phase of the SSVEP in addition to the frequency dimension [22]. The phase can encode information independently from the frequency. This feature can be used to multiply the number of available classes without impairing the frequency classification accuracy. However, technically, this requires synchronization between the stimulation hardware and the EEG signal recording.

A second approach to maximize the ITR by optimizing stimulation parameters is to enhance the classification accuracy TP by maximizing the signal to noise ratio (SNR) of the SSVEP. As the frequency dependence of the SSVEP response is subject-specific [8], the frequencies for encoding different target classes should be selected carefully. This also explains why the number of classes is hard to maximize: the more frequencies are employed to encode more classes, the lower the average SNR is ; thus, there is an optimum number of frequency classes per subject. In this study, we optimized the stimulation frequencies for each subject but used a fixed number of classes.

### B. Optimizing the Signal Processing

The attempt to maximize the SNR of the SSVEPs by optimizing stimulation parameters was paralleled by a corresponding effort in the development of the signal processing pipeline.

The first processing step consists of a single bandpass filter which selects a band containing the first and second harmonics of all SSVEP frequencies used.

As SSVEP responses are known to be largest over the visual cortex, only the posterior 14 of the 32 EEG channels were used. To this end, we employed a spatial filter based on a two-step application of the CCA, which we call CCA with partially fixed spatial filter (CCAFSF). Opposed to the standard CCA approach [14], which calculates new spatial filters based on the limited data each short live trial offers, we instead predetermined the spatial filters on large offline data and injected the spatial information, which we consider to be time independent, into the live feature extraction process.

The majority of classification methods employed for SSVEP BCIs assume that the data to be classified result from the subject gazing at any of the targets. A fixed number of samples is collected (typically lasting a few seconds), classified, and feedback about the result is given to the subject. Then, the cycle starts over by selecting the next target. We observed that transitory intervals (TIs) introduced by switching gaze from one target to the next result in noninformative periods in time which have a negative impact on the TP if they are included in the classification process. Thus, the length of the window of analysis ($l$) and the latency of the window of analysis with respect to the trial start ($\Delta t$) are critical parameters. A previous study [14] has shown that classification accuracy improves with the increasing window length. Longer input time requirements, however, also have a detrimental effect on the ITR [by increasing $\Delta t_{\mathrm{ITR}}$ in (1)]. Our aim is therefore to simultaneously optimize the time available for each target selection $L$, the window of analysis length $l$, and the offset of the window of analysis $\Delta t$. This optimization is performed for two classifier types (FE and TI), trained on data collected during both TIs as well as during periods of the fully entrained (FE) SSVEP.

### C. Optimizing the Mode of Operation

While performing continuous input to a SSVEP BCI the time the subject needs to select the next target, apparent latency of the SSVEP ([17], [23]), as well as the acquisition and processing of EEG data produce delays in the BCI's closed control loop (input, classification, cue for the next input). Due to these delays, each classification event is followed by a period of data containing the already classified SSVEP. This period is unavailable for classification of the last target and uninformative for the next.

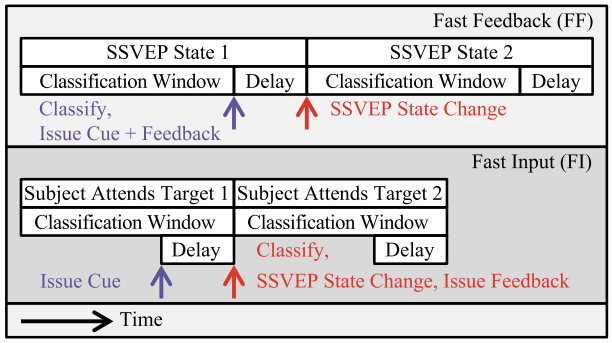


Fig. 1. Comparison of the temporal structure of FI and FF mode. Issuing the cue to switch targets results in a delayed change of the SSVEP. By issuing the cue early and independent from classification (FI mode), information in the "Delay" interval becomes available for classification.

The period directly adds to the loop's roundtrip time, decreases the time slots available for subject to system transmission, and represents a mayor negative impact on the system's ITR. We call the classical mode of the BCI operation, in which classification result feedback and cue to select the next target are represented by a singular event, FF mode (see Fig. 1, top panel), as it enables the subject to seemingly immediately react to the BCIs feedback. In the meantime the classifier discards uninformative, delayed data, while it waits for the next input to occur ("Delay" period in Fig. 1).

In anticipation of the delayed switch between SSVEP states, the cue for the next target can be issued early and independent of the classification and its feedback. By decoupling classification from cue and accurately anticipating the delay, all available time slots become available for subject-to-system transmission, increasing the bandwidth utilization and thus ITR. In this FI mode (see Fig. 1, bottom panel), however, the delays of the roundtrip are not negated. Instead, they are relocated but only relocated from the BCI's input (subject-to-system) to the BCIs output (system-to-subject) information streams. Both FF and FI mode are thus negatively affected by the delays of the system's roundtrip. In the FI mode, the early cue allows to exploit all available time slots for subject-to-system transmission at the cost of delayed classification and thus delayed feedback of the result. The FF mode allows to directly react to the classification result of the very last input at the cost of a lower bandwidth utilization and ITR.

Both FF and FI mode allow to react to feedback. In the FI mode, however, the subject may react only to the second to last, not the last input (or, depending on the proportion of the delays and the time allocated for each input, even older inputs). The correction of the second to last input was not tested here and is considered both more difficult to operate and less intuitive for the subject. Eventually, the preference for an FI or an FF mode will depend on the requirements imposed by the BCI's application.

In order to separate cue from feedback in the FI mode, we suggest to use an audio cue to pace the user selecting the next target, like a metronome paces a musical performance. In this schema, a continuous stream of input (like continuous typing on a keyboard) is sent to the BCI and is answered with a delayed but continuous stream of classification results (see Fig. 1).

In this study, the subjects did not receive classification result feedback. Instead, subjects only acted as driven by the classification result (which was neutral to the classification result). In the FF mode, this cue was bound to the same timing constraints (see above) as classification result feedback. The final ITR estimation therefore does not incorporate the additional delays expected from the reaction to a false classification result, but only represents the unidirectional subject-to-system information stream at the output of the Level 1 classifier [24].

## III. Experimental Setup and Data Analysis

### A. Hardware

Visual stimulation was generated by an array of 16 RGB LEDs, each with a conical diffusor lens of 3-cm diameter (visual angle of 5°) and calibrated to emit white light. Elements were arranged on the $4 \times 4$ array with a grid distance of 6 cm (10° visual angle). To the left, the LED array was extended by an additional column of four static fixation spots used as targets for the no-control state. Rectangular signals with a duty cycle of 50% were used to drive the LEDs as well as the EEG amplifiers trigger inputs. Sampling of the EEG and trigger signals as well as activation of the LED stimulators was driven by the same master clock. This is expected to minimize the SNR loss during the frequency analysis caused by slight deviations between the stimulation and sampling clocks.

A loudspeaker provided auditory cues to the subject. The corresponding trigger signal was recorded using a Schmitt trigger connected to the coil of the electromagnet driving the loudspeaker (see Fig. 2).

### B. EEG Recording

Seven healthy subjects aged between 17 and 45 (mean 28 years) were seated upright in a relaxed position on a chair in front of the LED array (distance 35 cm). EEG signals from the 14 electrodes closest to position Oz in the 10–20 system were recorded with a BioSemi 32 ActiveTwo amplifier at 1024-Hz sampling rate. The recording took place in a dimmed environment with some residual light coming from a static image on a 15-in TFT screen placed outside the field of view of the subject.

### C. Experimental Conditions and Tasks

To determine the parameters that maximize the ITR, we recorded data from four different experiments. Total recording duration was about 130 min per subject. Subjects were offered to take a break after each experiment. The first experiment was used to determine the SSVEP signal strengths at different stimulation frequencies and different feature extraction methods in order to select the four frequencies and corresponding feature extraction methods with the highest contrast. Data from the second experiment consisted of prolonged exposition to each of these four stimulation frequencies in four different phases each. They were used to train "FE classifiers" on the FE SSVEP and to determine one spatial filter per stimulus frequency. In

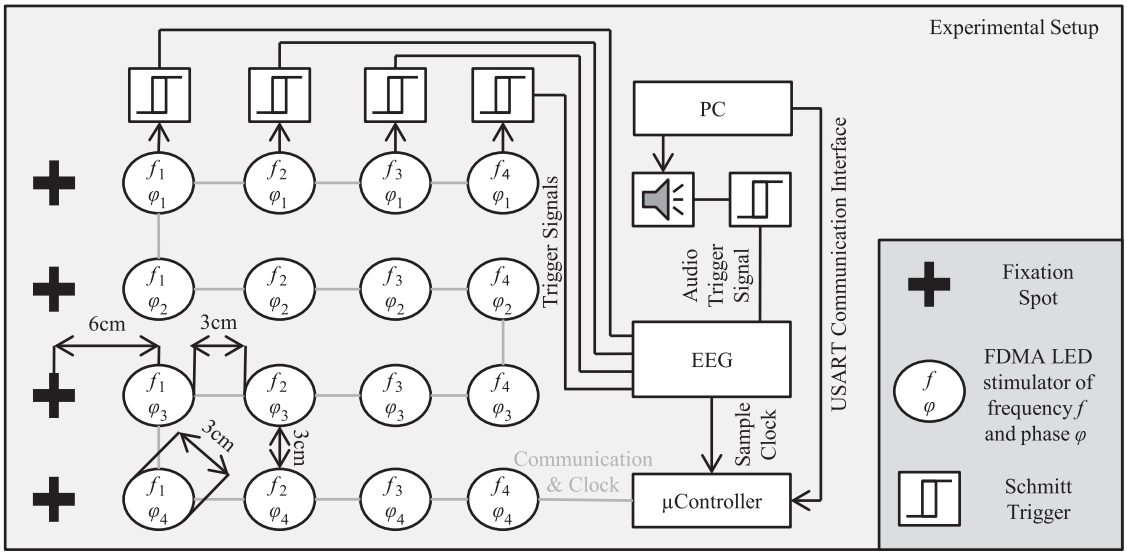


Fig. 2. Schematic overview of the stimulation and recording hardware setup.

the third experiment, we introduced transitions between steady states which occurred in fixed intervals of 3 s (triggered by an audio cue). The data were used to train "TI classifiers" for the time course of the SSVEP transitions and to estimate optimal input rates to be recorded in the fourth experiment using a heuristic. Using the fourth experiment's data, we fine-tuned temporal parameters for feature extraction and determined the ITR of the BCI under various modes of operation (FI or FF), with or without the no-control state.

***Experiment 1— Stimulus Parameter Optimization:*** In the first experiment, subjects were asked to overtly attend to each of the simultaneously presented targets on the LED array in a given sequence. Data of $\boldsymbol{L} = \mathbf{50}\ \boldsymbol{s}$ continuous exposure to each stimulus were recorded. After an optional period of rest for which the subjects did not have any instructions, they proceeded to the next target until each of them was attended once. The number of different stimulation frequencies $\boldsymbol{S}$ was set to 16. Stimulation frequencies were defined by their wavelengths $\boldsymbol{T_s}$ as an integer multiple of EEG samples. Furthermore, all wavelengths were restricted to four times an integer number of EEG samples, enabling accurate generation of four different phase angles per stimulation frequency. The 16 stimulation frequencies used in this experiment were $1024/[96\ 92\ 88\ 84\ 80\ 76\ 72\ 68\ 64\ 60\ 56\ 52\ 48\ 44\ 36\ 32]$ Hz.

***Experiment 2— Spatial Filters and FE Data:*** The second experiment introduced the no-control state and different phases of the stimulation frequencies. The 16 LEDs plus the four no-control state fixation spots were used as targets. The targets were selected in the same way as described for the first experiment. The LEDs were driven with $S = 4$ different frequencies $\hat{\boldsymbol{f}}_1, \hat{\boldsymbol{f}}_2, \ldots, \hat{\boldsymbol{f}}_S$ at four phase angles of 0°, 90°, 180°, and 270° each. Data of $\boldsymbol{L} = \mathbf{50}\boldsymbol{s}$ continuous exposure to each stimulus were recorded. The frequencies were selected to maximize the SSVEP detection accuracy of the respective subject based on results of the first experiment. One spatial filter per stimulation frequency was determined based on one partition of the data (see below), while the remaining data were used as training samples for the FE classifier.

***Experiment 3— Temporal Filters and TI Data:*** The third experiment used the same targets as the second experiment, but gathered data of the transition between targets instead of prolonged exposition to each target. This experiment consisted of four blocks, and each block consisted of four trials for each of the 20 targets (16 LEDs plus 4 fixation spots) in randomized order, with the constraint that the same target did not appear twice in a row. This resulted in 80 transitions per block. Each block was started by the subject via button press at his/her own discretion. The subject was instructed to switch gaze and attention to the next target as soon as a short auditory beep was perceived which was issued every $\boldsymbol{L} = \mathbf{3}\ \boldsymbol{s}$ and marked the beginning of each trial. Feedback about the last classification result was not given. In parallel to each currently attended target, the next target to be attended was visually indicated by changing its color from white to red. The color switching occurred simultaneously with the auditory cue. Using this cueing scheme, the subject always overtly attended a white flicker as the current target and covertly searched for a red flicker as the next. In the first trial, when there was no previous indication of which target to fixate, subjects were instructed to attend any of the white flicker stimuli. This first trial was omitted from further analysis. The visual cue for the four no-control state fixation spots was defined as none of the 16 LEDs being red. Subjects could freely choose which of the four spots to attend in this case.

***Experiment 4— Measuring ITR:*** In the fourth experiment, the time available for each input $\boldsymbol{L}$ was introduced as a parameter. One session with the otherwise same procedure and parameters as described for the third experiment (apart from $\boldsymbol{L}$) was recorded for each tested value of $\boldsymbol{L}$. Values for $\boldsymbol{L}$ started at the optimum predicted by the heuristic on data of the third experiment and were gradually decreased and increased in steps of $\mathbf{1}/\mathbf{8}\ \boldsymbol{s}$ to search for the FI mode ITR optimum. Some subjects reported that shorter settings of $\boldsymbol{L}$ were "too fast" in which case we stopped decreasing it.

### D. Signal Processing

***1) Preprocessing:*** EEG data were processed with a finite impulse response (FIR) minimum order equiripple bandpass filter with $-\mathbf{80}\boldsymbol{dB}$ attenuation in the stop bands and corner frequencies at the lowest presented stimulus frequency minus 2 Hz, and the highest stimulus frequency times 2 plus 2 Hz. The widths of the rising and falling slope of the filter were set to 5 Hz. This filter setting retained the first and second harmonics of every stimulation frequency in the pass band. The filter was restarted at the beginning of each block. All further processing of the data after filtering was done on data windows with a length $\boldsymbol{l}$ consisting of an integer number of bins with a fixed size of 1/8 s each. Bins containing the FIR startup time (length of the FIR kernel) were removed from further processing. Effected trials at the beginning of each block were omitted.

***2) Frequency Feature Extraction:*** We used CCA to determine the maximal correlation of the EEG signal with sine waves of the respective stimulation frequencies. In general, the CCA finds the maximal correlation between two multidimensional random variables in the matrices $\boldsymbol{X}$ ($\boldsymbol{d_1}$-by-$\boldsymbol{n}$) and $\boldsymbol{Y}$ ($\boldsymbol{d_2}$-by-$\boldsymbol{n}$), where $n$ is the number of samples, and $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$ are the number of variables or channels in $\boldsymbol{X}$, respectively, $\boldsymbol{Y}$. The CCA determines two $\boldsymbol{d_1}$-by-$\boldsymbol{d}$ and $\boldsymbol{d_2}$-by-$\boldsymbol{d}$ spatial filter matrices $\boldsymbol{W_x}$ and $\boldsymbol{W_y}$, where $\boldsymbol{d}$ is the minimum of $\boldsymbol{rank}(\boldsymbol{X})$ and $\boldsymbol{rank}(\boldsymbol{Y})$, which constitute linear combinations of the channels in $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, such that the $\boldsymbol{d}$-by-$\mathbf{1}$ correlation coefficients $\boldsymbol{R}$ between $\boldsymbol{x}$ (the $\boldsymbol{n}$-by-$\boldsymbol{d}$ projection of $\boldsymbol{X}$ on $\boldsymbol{W_x}$) and $\boldsymbol{y}$ (the $\boldsymbol{n}$-by-$\boldsymbol{d}$ projection of $\boldsymbol{Y}$ on ) are maximized. These projections are called canonical variants

$$x = X^T W_x \tag{2}$$

$$y = Y^T W_y. \tag{3}$$

In order to find the maximum $\boldsymbol{r}$ of the correlation coefficients $\boldsymbol{R}$, the CCA solves the following problem [14]:

$$r = \underset{W_x, W_y}{max} R(x,y) = \frac{E[x^T y]}{\sqrt{(E[x^T x]\, E[y^T y])}}. \tag{4}$$

In our case, $\boldsymbol{X}$ contained multiple channels of an EEG data window of analysis of length $\boldsymbol{l}$. The matrix $\boldsymbol{Y}$ was defined as a reference signal of length $\boldsymbol{l}$ for the sine wave to be detected in $\boldsymbol{X}$ by correlation. A set of reference signals $\boldsymbol{Y_s}$ corresponding to the stimulation frequencies $\boldsymbol{f_s}$ yielded scalar feature values $\boldsymbol{r_s}$ for each frequency. They were combined in the feature vector $\boldsymbol{F}$ used for classification

$$F = \left( r_1\ r_2\ \ldots\ r_S \right). \tag{5}$$

The reference signals $\boldsymbol{Y_s}$ model the respective SSVEP response by a linear combination of sine and cosine functions

$$\mathrm{Y_{s,h}} = \begin{pmatrix} \sin(2\pi\ \mathrm{f_s}\ \mathrm{t}) \\ \cos(2\pi\ \mathrm{f_s}\ \mathrm{t}) \\ \cdots \\ \cdots \\ \sin(2\pi\ \mathrm{h}\ \mathrm{f_s}\ \mathrm{t}) \\ \cos(2\pi\ \mathrm{h}\ \mathrm{f_s}\ \mathrm{t}) \end{pmatrix}. \tag{6}$$

Here, $\boldsymbol{h}$ is the number of harmonics $\boldsymbol{h} \in [\mathbf{1}, \mathbf{2}]$, i.e., $\boldsymbol{Y_{s,1}}$ consists of one pair of sine and cosine functions approximating the first harmonic of the SSVEP, and $\boldsymbol{Y_{s,2}}$ consists of an additional pair of sinusoids in order to model the second harmonic as well. By adjusting the weights in $\boldsymbol{W_y}$, the CCA is able to fit the phase and the amplitude of the sinusoids in $\boldsymbol{y}$ to the data in $\boldsymbol{x}$ [compare (7)]. The filter weights in $\boldsymbol{W_y}$ thus constitute the basis of the time dependent information required for the feature values

$$a\sin(\alpha) + b\cos(\alpha) = c\sin(\alpha+\varphi)$$
$$c = \sqrt{a^2+b^2};\ \varphi = \tan^{-1}\left(\frac{a}{b}\right). \tag{7}$$

***3) Spatial Filtering Using CCAFSF:*** The weights of $\boldsymbol{W_x}$, however, contribute spatial filters for the EEG sensor space information which we hypothesized to be time independent. Instead of recalculating $\boldsymbol{W_x}$ for each short live trial (as done by the standard CCA [14]), $\boldsymbol{W_x}$ can be precalculated with increased precision by using large amounts of offline data. Once determined, the spatial information can be injected into the live process operating on short data windows without compromising the BCIs input rates.

For each $\boldsymbol{f_s}$, we recorded (in Experiment 2) a single long FE SSVEP trial $\hat{\boldsymbol{X}}_{\boldsymbol{s}}$. A first CCA was applied to each $\hat{\boldsymbol{X}}_{\boldsymbol{s}}$ and the corresponding $\hat{\boldsymbol{Y}}_{\boldsymbol{s}}$ to train the frequency specific spatial filters $\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$. This also yielded $\hat{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\hat{\boldsymbol{y}}_{\boldsymbol{s}}$. In subsequent experiments, short trials $\boldsymbol{X}$ were projected onto each $\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$, resulting in the canonical variants $\tilde{\boldsymbol{x}}_{\boldsymbol{s}}$

$$\hat{x}_s = \hat{X}_s^T \hat{W}_{x_s} \tag{8}$$

$$\hat{y}_s = \hat{Y}_s^T \hat{W}_{y_s} \tag{9}$$

$$\tilde{x}_s = X^T \hat{W}_{x_s}. \tag{10}$$

One feature value per $\boldsymbol{f_s}$ was extracted using the frequency-specific projection $\tilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and the corresponding $\boldsymbol{Y_s}$ as input to a second standard CCA step. The resulting features were grouped in $\boldsymbol{F_v}$ for classification

$$F_v = \left( r_{v,1}\ r_{v,2}\ \ldots\ r_{v,S} \right). \tag{11}$$

By using $\tilde{\boldsymbol{x}}_{\boldsymbol{s}}$ instead of $\boldsymbol{X}$, the degree of freedom of the error prone second CCA's spatial filter estimate was thus constrained by the spatial information contained in $\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$. This successive application of two standard CCAs [14] to precalculate time-independent properties of the data (location of the SSVEP generating source), and separate them from time-dependent ones (phase and amplitude of each $\boldsymbol{f_s}$) is called CCAFSF and is depicted as a flowchart (see Fig. 3). CCAFSF is considered the state-of-the-art approach for CCA SSVEP frequency feature extraction increasing feature contrast and classification accuracy.

The channel numbers which yielded the maximum correlation coefficient for each individual $\boldsymbol{f_s}$ during spatial filter ($\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$) training were stored as vector $\hat{\boldsymbol{c}}_{\boldsymbol{s}}$ and used during phase feature extraction

$$\hat{c}_s = \underset{\hat{W}_{\hat{x}_s}, \hat{W}_{\hat{y}_s}}{argmax} \hat{R}_s(\hat{x}_s, \hat{y}_s). \tag{12}$$

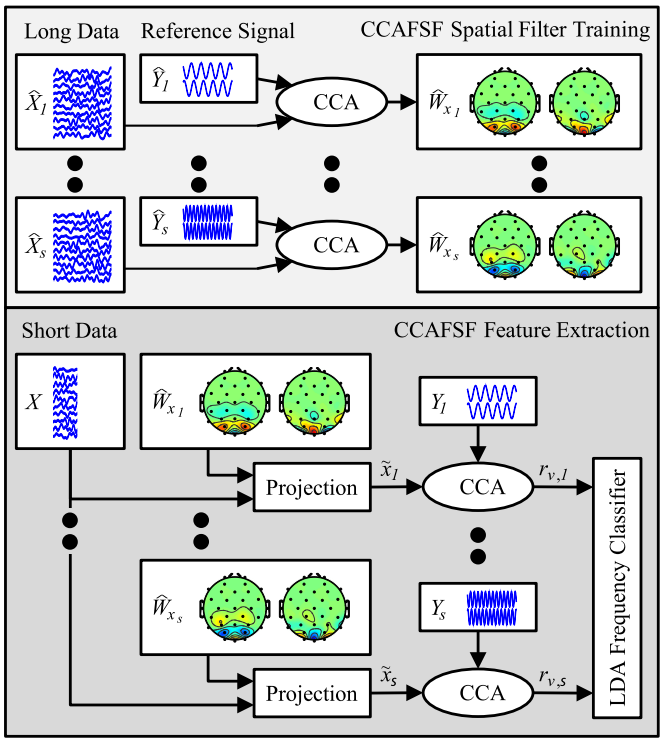


Fig. 3. Schematic overview of CCAFSF spatial filter training (top) and frequency feature extraction (bottom). For each frequency $\mathrm{f_s}$, a single long trial $\hat{\mathrm{X}}_\mathrm{s}$ and a reference signal $\hat{\mathrm{Y}}_\mathrm{s}$ with a frequency specific, optimized number of harmonics ($\mathrm{h} \stackrel{\mathrm{def}}{=} 1$ for purposes of this example) was used as input to a first CCA to train the frequency specific spatial filters $\hat{\mathrm{W}}_{\mathrm{x_s}}$. Depending on the $\mathrm{h_s}$ used for the individual $\hat{\mathrm{Y}}_\mathrm{s}$, the resulting $\hat{\mathrm{W}}_{\mathrm{x_s}}$ consisted of either 2 ($\mathrm{h} = 1$) or 4 ($\mathrm{h} = 2$) spatial filter vectors (depicted as plots of the topography). In the second step (bottom), all $\hat{\mathrm{W}}_{\mathrm{x_s}}$ were applied to each short live trial (of unknown frequency class) before a second CCA was used to extract one feature value ($\mathrm{r}_{v,s}$) from each frequency specific projection $\tilde{x}_s$.

***4) Phase Feature Extraction:*** Fourier coefficients for each $\boldsymbol{f_s}$ were computed by a standard FFT on a single- frequency specific data channel. The real and imaginary parts of the Fourier coefficients were combined in a phase feature vector $\boldsymbol{P_s}$ used for classification.

Let $\boldsymbol{g_s}$ be a time-domain input data vector containing $\boldsymbol{N}$ samples. For each $\boldsymbol{f_s}$, we define its wave length $\boldsymbol{T_s}$ and the maximum number of samples $\boldsymbol{N_{\mathrm{FFT},s}}$ to be used for the FFT which result in a frequency-domain bin $\boldsymbol{k_s}$ that closely matches the desired frequency. Here, the use of the same clock signal for the sampling and stimulus generation process allowed for an exact match

$$T_s = \frac{1}{f_s}; N_{fft,s} = \left[\frac{N}{T_s}\right] T_s; k_s = \frac{N_{fft,s}}{T_s}$$

$$G_s = \frac{1}{N} \sum_{n=0}^{N_{fft,s}} g_s(n)\, e^{\frac{-2i\pi k_s * (n-1)}{N_{fft,s}}}$$

$$P_s = \begin{pmatrix} \Re(G_s) & \Im(G_s) \end{pmatrix}. \tag{13}$$

The first rising trigger edge of the four LEDs running at $0^\circ$ phase was used as a phase reference per corresponding data window. Depending on the frequency of interest and the method used for frequency feature extraction (CCA or CCAFSF), a different data channel was used as input vector $\boldsymbol{g_s}$ to the same FFT (13). In the CCA case, we used the channel Oz as $\boldsymbol{g_s}$ for every frequency. In the CCAFSF case, we first projected the sensor space data $\boldsymbol{X}$ on the frequency-specific spatial filter matrix $\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$, which resulted in the projection $\hat{\boldsymbol{W}}_{\boldsymbol{x_s}}$ (10). From $\tilde{\boldsymbol{x}}_{\boldsymbol{s}}$, we selected the single channel $\hat{\boldsymbol{c}}_{\boldsymbol{s}}$ as input $\boldsymbol{g_s}$ to the FFT (13). The resulting phase feature vector was denoted $\boldsymbol{P_s}$ in the CCA case and $\boldsymbol{P_{v,s}}$ in the CCAFSF case.

***5) DQDA Classifier Cascade:*** Throughout all experiments, trials were classified using the same cascade of two diagonal quadratic discriminant analysis (DQDA) classifiers. The first one distinguished the control states from the no-control state using the feature vectors $\boldsymbol{F}$ (CCA) or $\boldsymbol{F_v}$ (CCAFSF). If a control state was detected, the second classifier discriminated between the four phase classes of the detected frequency using the phase features $\boldsymbol{P_s}$ (CCA) or $\boldsymbol{P_{v,s}}$ (CCAFSF).

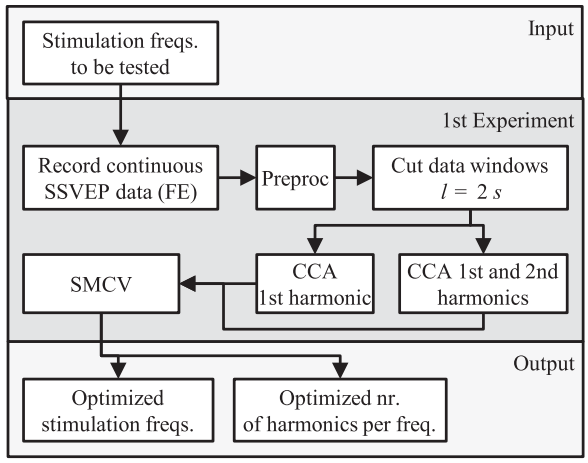


Fig. 4. Schematic overview of the first experiment.

## E. Parameter Optimization

***1) Stimulation Frequency Optimization Using SMCV:*** The standardized mean of a contrast variable (SMCV) [25] method was applied to data of the first experiment to determine the optimal number of harmonics (1 or 2) to be used in the CCA reference signals of each frequency and to find those four stimulation frequencies [8] which maximize group difference (see Fig. 4).

Let there be S random variables $\boldsymbol{H_1}, \boldsymbol{H_2}, \ldots, \boldsymbol{H_S}$ with the mean values $\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \ldots, \boldsymbol{\mu_S}$. A contrast variable $\boldsymbol{C}$ is defined by

$$C = \sum_{i=1}^{S} w_i H_i. \tag{14}$$

where $\boldsymbol{w_i}$'s are weights for the groups which satisfy

$$\sum_{i=1}^{S} w_i = 0. \tag{15}$$

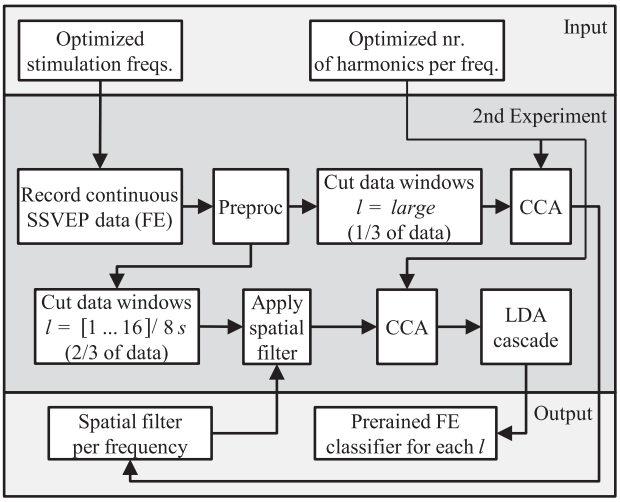


Fig. 5. Schematic overview of the second experiment.

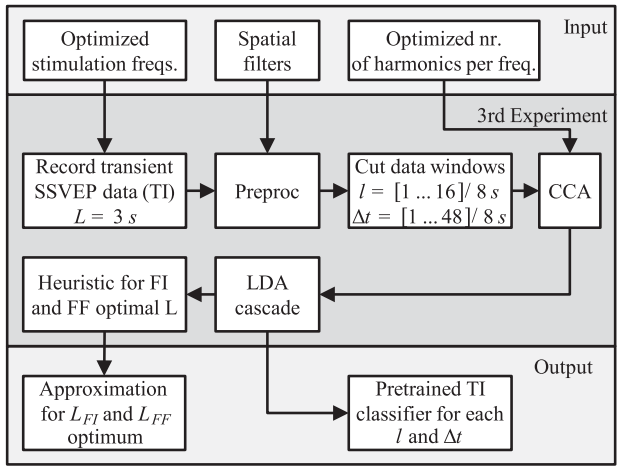


Fig. 6. Schematic overview of the third experiment.

The SMCV for $\boldsymbol{C}$ denoted by $\boldsymbol{\lambda}$ is defined [25] as

$$\lambda = \frac{E(C)}{stdev(C)} = \frac{\sum_{i=1}^{S} w_i \mu_i}{\sqrt{Var\left(\sum_{i=1}^{S} w_i H_i\right)}}. \qquad (16)$$

Larger differences in mean values and smaller variances of each contrasted group result in larger $\lambda$ values.

Trials were cut from each $50s$ data segment of the first experiment as windows with a length $\boldsymbol{l = 2\ s}$ and $\boldsymbol{l - 1/8\ s}$ overlap. For any one combination of the parameters frequency and number of harmonics, we extracted features from all trials using the corresponding CCA reference signal $\boldsymbol{Y_{s,h}}$. Feature values of those trials during which the current frequency of interest was attended were grouped in $\boldsymbol{H_1}$ and weighted by $\boldsymbol{w_1 = 1}$. Feature values of all other trials were grouped by their respective attended frequency ($\boldsymbol{H_2, \ldots, H_S}$) and weighted by $\boldsymbol{w_2, \ldots, w_S = -1/(S-1)}$ to satisfy (15). Repetition of this process for all tested frequencies and number of harmonics yielded one $\boldsymbol{\lambda_{s,h}}$ for each combination.

We first found the maximum $\boldsymbol{\lambda_{s,h}}$ among the number of harmonics for each individual frequency and then determined the four largest remaining $\boldsymbol{\lambda_{s,h}}$ across frequencies to maximize group separability. The resulting four stimulation frequencies and corresponding number of harmonics were considered as optimal and used exclusively in all subsequent experiments.

In order to estimate the variance in the reported $\boldsymbol{\lambda_{s,h}}$ values, we repeatedly performed their calculation on randomized partitions each containing only 90% of the data (delete-10% jackknife).

***2) Continuous SSVEP:*** In the second experiment, we recorded continuous ($\boldsymbol{50\ s}$) exposition to the no-control state targets and four stimulation frequencies in four phases each. Data were used to train the spatial filters $\boldsymbol{\hat{W}_{x_s}}$ for CCAFSF and as training datasets for the "FE classifiers." We also used the data to report performance differences of CCA versus CCAFSF (see Fig. 5).

The first one-third of the $\boldsymbol{50\ s}$ exposition to each $\boldsymbol{f_s}$ at 0° phase was used as $\boldsymbol{\hat{X}_s}$ to train the corresponding spatial filters $\boldsymbol{\hat{W}_{x_s}}$ to be used in subsequent experiments.

The remaining two-third of data was cut to data windows of length $\boldsymbol{l = [1, 2, \ldots, 16]/8\ s}$ with an overlap of $\boldsymbol{l - 1/8\ s}$. Frequency and phase features were extracted by CCA and CCAFSF from each data window. The features were grouped by the feature extraction method and $\boldsymbol{l}$. Each group defined an FE classifier training set to be used in subsequent experiments.

In order to report performance differences between CCA and CCAFSF, the features of each group were also classified by the DQDA cascade. We separated each group in four nonoverlapping sets and used them in all possible combinations as training (two sets) and test (two sets). For each permutation of the sets, we used 100 randomized partitions each containing 90% of the training and test data (delete-10% jackknife). Virtual ITRs (VITR) were calculated based on the resulting TP rates using (1) and $\boldsymbol{\Delta t_{ITR} = l}$. We call this ITR "virtual" because we defined the time required to transmit each character $\boldsymbol{\Delta t_{ITR}}$ as the length of data $\boldsymbol{l}$ used for classification instead of the time available for each target selection $\boldsymbol{L}$. The second experiment was specifically designed to not contain any TIs and thus no switches between targets. Consequently, there was no real transfer of information in this experiment.

***3) Transitions Between SSVEPs:*** The TI data from the third experiment was used to train "TI classifiers" in analogy to how the data from the second experiment was used to train the FE classifiers, and to estimate the optimal time between switching targets $\boldsymbol{L}$ for the fourth experiment (see Fig. 6).

The parameter $\boldsymbol{\Delta t}$ was introduced as the offset of the right-side corner of each windows of analysis relative to the audio trigger marking the beginning of each trial. Windows of analysis were cut from each trial for each combination of the parameters $\boldsymbol{l = [1, 2, \ldots, 16]/8\ s}$ and $\boldsymbol{\Delta t = [1, 2, \ldots, 48]/8\ s}$. Features were extracted from each window by CCAFSF and grouped by parameters values. Each group was used as a training set for one TI classifier.

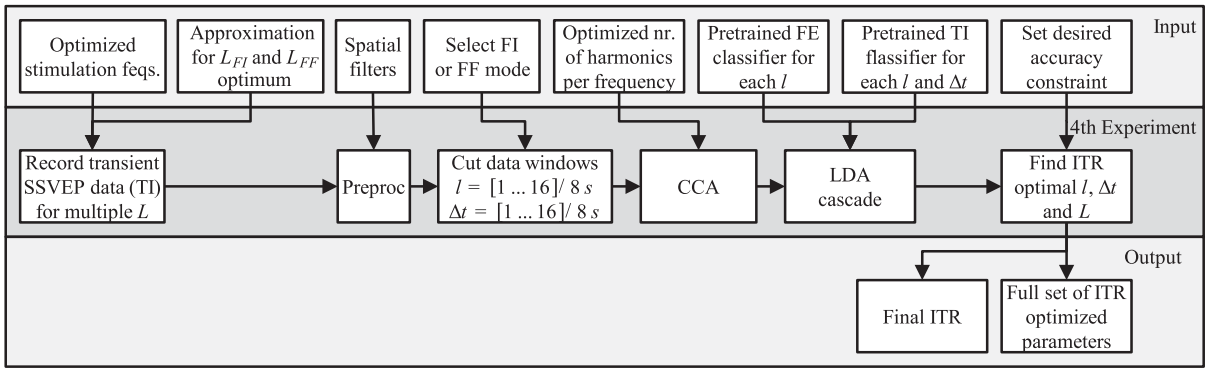


Fig. 7. Schematic overview of the fourth experiment.

All groups were also classified by the FE classifiers trained on the same parameter values ($l$, CCAFSF). To estimate the variance of the TP, we used 100 iterations of repeated random subsampling using 90% of the training FE and test TI data. Optimal parameters were determined by finding the ITR maximum when using the above TP rates, $\boldsymbol{\Delta t_{ITR} = L}$, and (1).

The time between switching targets $\boldsymbol{L}$ was thus far not considered as a parameter. Instead of recording and testing a wide range of $\boldsymbol{L}$ values, we roughly approximated the ITR optimal $\boldsymbol{L}$ for FI and FF mode based on the above $\boldsymbol{L = 3\ s}$ data TP rates. VITR values were determined for every combination of the parameters $\boldsymbol{l}$ and $\boldsymbol{\Delta t}$ using (1), the above TP rates, and the estimated minimum input time requirement for the respective mode ( $\boldsymbol{\Delta t_{ITR} = L_{FI}}$ for FI or $\boldsymbol{\Delta t_{ITR} = L_{FF}}$ for the FF mode). $\boldsymbol{L_{FI}}$ and $\boldsymbol{L_{FF}}$ were each modeled as functions of $\boldsymbol{\Delta t}$ (see below). The $\boldsymbol{L_{FI}}$ value associated with the VITR($\boldsymbol{l}$, $\boldsymbol{\Delta t}$) maximum was considered optimal and used as initial value for $\boldsymbol{L}$ in the fourth experiment.

The classification in the FF mode and simultaneous issuing of the audio cue at $\boldsymbol{\Delta t}$ resulted in lagged reception of the corresponding audio trigger signal ($\boldsymbol{162 \pm 26\ ms}$ ceiled to $\boldsymbol{t_{lag} = \frac{2}{8} s = 250\ ms}$ due to the resolution of the analysis time bins). In the FF mode, the minimum required time between switching targets while classifying at $\boldsymbol{\Delta t}$ was thus modeled as

$$\boldsymbol{L_{FF}\ (\Delta t) = \Delta t + t_{lag}.} \tag{17}$$

In the FF mode, the interval of length $\boldsymbol{t_{lag}}$ between classification and the start of the next trial cannot be used effectively for classification, but it adds to the time required for each input.

To access the information in this interval, classification and cue were decoupled in time in the FI mode. We assumed that the cue can be issued $\boldsymbol{t_{lag}}$ earlier while still classifying with comparable TP rates at $\boldsymbol{\Delta t}$. Furthermore, the subject's reaction time to each cue signal was assumed to result in additional delays (switching of overt attention, latency of the visual pathway [23]) of the SSVEP response which were estimated at $\boldsymbol{t_{react} = 250\ ms}$. Until switching occurred, the subject would still maintain the last SSVEP state. The minimum target selection time required in the FI mode was thus modeled as

$$\begin{aligned} \boldsymbol{L_{FI}\ (\Delta t)} &= \boldsymbol{L_{FF}\ (\Delta t) - t_{react} - t_{lag}} \\ &= \boldsymbol{\Delta t - t_{react}.} \end{aligned} \tag{18}$$

***4) Application of the BCI:*** Based on the data from the fourth experiment, we performed an exhaustive search in the vicinity of the estimated optimal trial length $\boldsymbol{L_{FI}}$ in order to find subject-specific settings for $\boldsymbol{L}$, $\boldsymbol{l}$,$\boldsymbol{\Delta t}$, and the type of classifier (FE or TI) which maximized the ITR of the simulated online BCI system in the FI mode (see Fig. 7). Features extracted using a particular parameter combination ($\boldsymbol{l}$,$\boldsymbol{\Delta t}$) were classified by those classifiers which had been trained on the exact same parameter setting. For comparison, parameters were also optimized for the FF mode. We did not, however, specifically vary $\boldsymbol{L}$ in the vicinity of $\boldsymbol{L_{FF}}$. Instead the data recorded in the vicinity of $\boldsymbol{L_{FI}}$ was used. Instead of performing online classification of the fourth experiment, only software trigger signals marking the point in time when the audio cue was issued from software (when to classify offline) were written online. Those trigger signals allowed to estimate the reported ITR performance distributions offline by twofold cross-validation (100 repetitions).

## IV. Results

### A. First Experiment—Stimulation Frequency Optimization

Data from the first experiment were used to determine four optimal subject specific stimulation frequencies $f_s$ and a corresponding number of harmonics for the CCA reference signals for each of them (summarized in Table I). The optimization was based on the size of effect values $\lambda_{s,h}$ calculated by SMCV. Whereas, the study in [14] did not find a significant impact of the number of harmonics $h$ on the classification performance, our data show that the SMCVs can be significantly increased for most subjects and frequencies by including the second harmonic. Using a mixed number of harmonics for the individual frequencies were found to be optimal for two out of seven subjects (cf., Table I).

### B. Second Experiment—CCAFSF and FE State analysis

Data from the second experiment were obtained during prolonged exposition to each of the four stimulation frequencies in four phases each. The resulting long FE data segments allowed determination of CCAFSF spatial filters and training of FE classifiers for simulated online classification of the fourth experiment.

TABLE I
SUBJECT STIMULATION PARAMETERS OVERVIEW

| | Stimulus $f_1$ | | | Stimulus $f_2$ | | | Stimulus $f_3$ | | | Stimulus $f_4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject No.** | **T [1/1024 s]** | **λ (Ø±std)** | **h** | **T [1/1024 s]** | **λ (Ø±std)** | **h** | **T [1/1024 s]** | **λ (Ø±std)** | **h** | **T [1/1024 s]** | **λ (Ø±std)** | **h** |
| **1** | 68 | 3.2±0.14 | 2 | 64 | 3.7±0.18 | 2 | 60 | 3.1±0.11 | 2 | 56 | 4.2±0.15 | 2 |
| **2** | 76 | 10.5±0.37 | 2 | 72 | 12±0.68 | 2 | 68 | 9.6±0.52 | 2 | 60 | 10.1±0.39 | 2 |
| **3** | 52 | 8.1±0.34 | 2 | 48 | 9±0.34 | 1 | 44 | 7.7±0.3 | 2 | 36 | 8.2±0.4 | 2 |
| **4** | 68 | 2.6±0.12 | 2 | 64 | 2.8±0.13 | 2 | 60 | 2.7±0.11 | 2 | 56 | 2.6±0.1 | 2 |
| **5** | 96 | 9.4±0.38 | 2 | 92 | 9.3±0.31 | 2 | 88 | 8.2±0.28 | 2 | 84 | 8.4±0.34 | 2 |
| **6** | 76 | 4.8±0.22 | 2 | 72 | 4.9±0.25 | 2 | 68 | 4.9±0.21 | 2 | 60 | 5.1±0.22 | 2 |
| **7** | 60 | 7.9±0.26 | 1 | 52 | 7.2±0.4 | 2 | 44 | 7.8±0.4 | 2 | 36 | 9.3±0.3 | 2 |

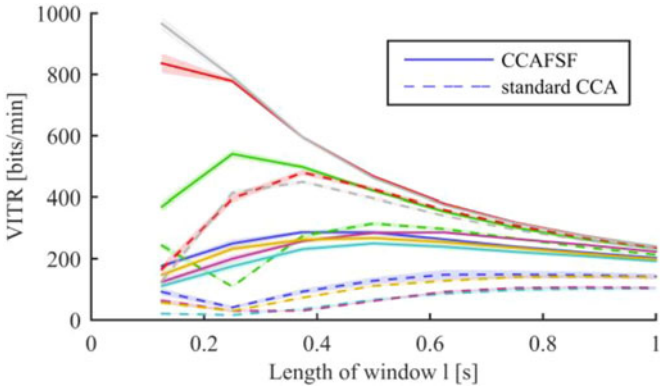


Fig. 8. VITR comparison of the CCA and CCAFSF algorithms on FE data (no transitions between classes; no loss of bandwidth caused by transitions) classified by FE classifiers over $l$ for each subject (color coded cf., Table I).

TABLE II
EFFECT OF CCAFSF ON FE DATA

| **Subject No.** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|
| **Maximal avg. VITR over *l* using CCA [Bits/min]** | 149 | 315 | 480 | 103 | 107 | 141 | 449 |
| **Maximal avg. VITR over *l* using CCAFSF [Bits/min]** | 287 | 540 | 837 | 250 | 286 | 268 | 966 |
| **VITR increase by CCAFSF [%]** | 93 | 72 | 74 | 144 | 167 | 89 | 115 |

We compared the efficiency of CCA and CCAFSF by extracting and classifying feature values from the FE using both methods. The resulting VITRs show an improved average classification accuracy and lower variance (see Fig. 8) in all subjects using CCAFSF. Classification of data from the FE SSVEP containing no transitions between targets (no real transfer of information; no loss of bandwidth caused by the transitions) yielded VITRs of up to $\sim 1\ \mathrm{kb/min}$. The respective maxima represent optimal tradeoffs between accuracy and the length of the window of analysis $l$. Other subjects seemed not to reach a global maximum in the range of considered window lengths ($l = [0.125 \ldots 1]\ s$). Using CCA, some subjects exhibited a nonmonotonic change of VITR with increasing $l$. Table II lists the FE VITR maxima for the methods CCA and CCAFSF as well as the relative improvement by CCAFSF for each subject based on the data of Fig. 8.

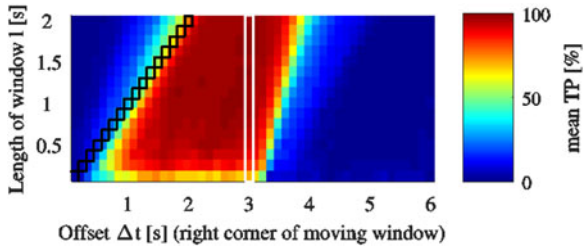


Fig. 9. Mean classification accuracies (color coded) of the FE classifier applied to TI data over $l$ and $\Delta t$ for one exemplary subject (No. 7). Displayed are two consecutive trials (length $L = 3s$). Color shows classification accuracies of the first trial. White boxes mark the first time bin of the second trial. Black boxes mark the first offsets which do not result in an overlap of the data window with the previous trial.

### C. Third Experiment—Transitions Between SSVEPs

The third experiment added information about the SSVEPs development in time by introducing transitions between the steady states. The data were used to train TI classifiers for simulated online classification of the fourth experiment.

FE classifiers were tested on the TI data for different parameter values of $l$ and $\Delta t$. The resulting TP rates (see Fig. 9) reveal a delayed start and end of the SSVEP w.r.t. the auditory cues. The SSVEP responses were also shown to be shorter than the exposition to the stimulus, suggesting the presence of a dead time. A shallower rising than falling slope (over $\Delta t$) of the TP rates indicated an initial period of SSVEP entrainment.

The $L = 3s$ TI data were further used to approximate the optimal trial length for FI and FF mode ($L_{\mathrm{FI}}$ and $L_{\mathrm{FF}}$) by determining the VITR maximum over $l$ and $\Delta t$. The VITRs displayed in Fig. 10 were calculated based on the accuracies shown in Fig. 9, using (1) and (18), as well as $\Delta \mathrm{t}_{\mathrm{ITR}} = \mathrm{L}_{\mathrm{FI}}$. The data show that with increased input rates, small window lengths, and early classification, higher ITRs can be expected at the expense of accuracy and variability. Table III lists the parameter values at which the individual VITR maxima were located.

### D. Fourth Experiment—Application of the BCI

The analysis of the fourth experiment used the information of all previous experiments, added the time available for each input $L$ as a parameter, and optimized $L$, $\Delta t$, $l$, and the type of classifier (TI or FE) by an exhaustive search for each subject. The resulting data revealed a successful reduction of the broad

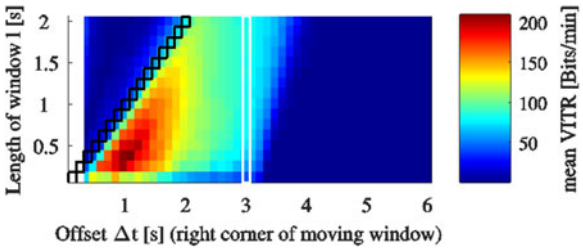


Fig. 10. Mean VITR values (color coded) over $l$ and $\Delta t$ calculated using the ITR (1), the classification accuracies shown in Fig. 9, and an estimate for the minimum required input time which results in comparable accuracies at each individual bin [(18) and $\Delta t_{\mathrm{ITR}} = L_{\mathrm{FI}}$ ]. The maximum VITR is located at $\Delta t = 1.125$ s and $l = 0.375$ s and corresponds to an estimated optimal FI mode trial length of $L_{\mathrm{FI}} = 0.875$ s.

TABLE III
STIMULATION LENGTH HEURISTIC PARAMETERS

| Subject No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\Delta t$ [s] | 1.25 | 1.375 | 1.25 | 1.25 | 1.5 | 1.125 | 1.125 |
| $l$ [s] | 0.75 | 0.5 | 0.375 | 0.75 | 0.875 | 0.625 | 0.375 |
| $L_{FI}$[s] | 1.0 | 1.125 | 1.0 | 1.0 | 1.25 | 0.875 | 0.875 |

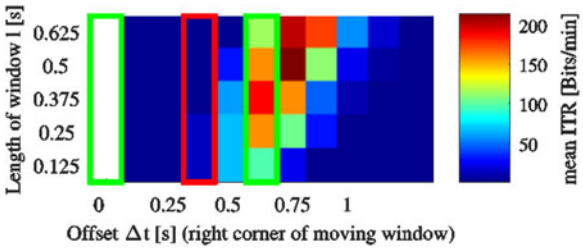


Fig. 11. Mean ITR over $l$ and $\Delta t$ for an exemplary subject (No. 7) using the optimal trial length $L = 625$ ms and classifier (TI). Two consecutive trials are displayed. Red outlines mark the time bin when the audio cue for the second trial was issued from software. Green outlines mark the reception of the audio cue trigger (for the first and second trial) by software.

plateaus of high classification accuracy caused by suboptimal parameter settings (see Fig. 9) to narrow peaks as illustrated by the example inFig. 11. Operating the BCI continuously with this optimal yet short time window required stable reaction times from the subject.

Overlaying the ITR($l$, $\Delta t$) in Fig. 11 with issuance (red rectangle) and reception (green rectangles) of the audio cue signal revealed time dependencies of both modes (FI and FF). Due to lags of the system ($162 \pm 26$ ms) and the resolution of the analysis (125 ms), the start of each trial (green rectangles) was located in the time bin $t_{\mathrm{lag}} = 250$ ms after issuing its cue from software. Furthermore, subtracting the optimal FI mode length of the window of analysis from its optimal right side offset and averaging over all subjects revealed an average interval of 357 ms between trial start and start of the SSVEP (cf., Table IV). The optimal time for classification was thus located in the consecutive trial. In the FF mode, the last opportunity to classify was the moment when the next cue (feedback) needed to be issued (red rectangle). Thus, the time bins [500 ms ... 1125 ms] (see Fig. 11) featuring high ITRs were inaccessible in the FF mode. The FI mode, however, allowed to exploit those high ITRs by classifying at any offset $\Delta t$, even after the next cue was already issued. This supports our idea that overlaying classification of the current trial with stimulation of the next can improve the ITR substantially.

By subtracting the optimal FF mode length of data window from the respective optimal length of trial and averaging over all subjects, we found that only $\sim 49.8\%$ of all available time slots were informative for classification. In the FI mode, while allowing for classification during the consecutive trial, the same parameter revealed that $\sim 22.5\%$ (205 ms on average cf., Table IV) of all available time slots were still uninformative for classification—a dead time of the SSVEP response.

Comparing FI and FF ITRs over all tested trial lengths (see Fig. 12), we found that the impact of the noninformative delays on the FF ITR increased with decreasing trial length. For short trials, the FF ITRs dropped toward 0 Bits/min as the time was completely consumed by the delays and dead time of the SSVEP response. FI ITRs approached FF ITRs asymptotically as the proportion of the uninformative overhead diminished with longer trials.

To lessen the negative effect of the SSVEP transients, we trained a set of TI classifiers on the transients. In our data, subjects with comparatively high ITR performance benefitted from using the TI classifier, whereas subjects with lower performance did not show large differences between classifier types (see Fig. 13).

Omitting the no control state class led to an ITR improvement for most subjects. The final system reached an average simulated online ITR of 181 Bits/min (peak value $295 \pm 12$ Bits/min) in the FI mode and an average ITR of 124 Bit/min (peak value $187 \pm 7$ Bit/min) in the FF mode. Table IV summarizes the final results and optimal parameter values (mode in case of $L$, $l$, and $\Delta t$).

During recording, the trial length was varied with the intention to optimize FI (but not FF) mode ITRs. In the FF mode, the optimal trial lengths equaled the largest value tested for three out of seven subjects, suggesting that not for every subject a global optimum was reached.

## V. DISCUSSION

The main rationale of our approach to maximizing the ITR of an FDMA SSVEP BCI was to divide it into its main components, to study their properties, and to optimize their parameters individually.

We started with the hardware components. The communication pathway of a FDMA SSVEP BCI has certain similarities with digital communication: the stimulator acts as the sender or generator of a number of different carrier frequencies. These are selected or modulated by the intent of the subject, received by the EEG sampling process, and interpreted by the analysis algorithm. Such communication benefits from a stable clock signal known to both sender and receiver. We designed a real-time-capable LED stimulator driven by the clock signal of the EEG amplifiers analog to digital converter. The operation based on this common, comparatively fast, and system-wide clock signal

TABLE IV
SUBJECT PERFORMANCE OVERVIEW (CCAFSF)

| Mode | Subject No. | ITR Ø ± std [Bits/min] | | Accuracy [%] | | Optimal classifier | | Optimal length of trial $L[s]$ | | Optimal length of data window $l$ [s] | | Optimal offset $\Delta t$ [s] | | $L - l$ [s] | | $\Delta t - l$ [s] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FI | 1 | 70±7 | 67±8 | 61.8±3.5 | 61.6±4.2 | TI | TI | 1.375 | 1.375 | 1.25 | 1.25 | 1.625 | 1.625 | 0.125 | 0.125 | 0.375 | 0.375 |
| | 2 | 146±7 | 167±9 | 88.5±1.6 | 95.2±2.0 | TI | TI | 1.125 | 1.125 | 0.75 | 0.875 | 1.25 | 1.25 | 0.375 | 0.25 | 0.5 | 0.375 |
| | 3 | 278±13 | 295±12 | 95.5±1.2 | 96.0±1.4 | TI | TI | 0.75 | 0.75 | 0.5 | 0.5 | 0.875 | 0.875 | 0.25 | 0.25 | 0.375 | 0.375 |
| | 4 | 127±9 | 138±10 | 81.7±2.7 | 86.5±2.7 | FE | TI | 1.25 | 1.25 | 1.0 | 1.0 | 1.375 | 1.375 | 0.25 | 0.25 | 0.375 | 0.375 |
| | 5 | 159±11 | 187±13 | 91.0±1.6 | 93.4±1.6 | FE | TI | 0.875 | 0.875 | 0.75 | 0.75 | 1.25 | 1.25 | 0.125 | 0.125 | 0.5 | 0.5 |
| | 6 | 187±12 | 181±13 | 93.5±1.7 | 96.6±1.4 | FE | FE | 0.75 | 0.75 | 0.5 | 0.5 | 0.625 | 0.75 | 0.25 | 0.25 | 0.125 | 0.25 |
| | 7 | 198±16 | 232±15 | 93.2±2.0 | 95.9±1.4 | TI | TI | 0.625 | 0.625 | 0.5 | 0.5 | 0.75 | 0.75 | 0.125 | 0.125 | 0.25 | 0.25 |
| FF | 1 | 39±6 | 46±7 | 46.1±3.3 | 50.1±4.1 | TI | TI | 1.375 | 1.375 | 0.75 | 0.75 | 1.375 | 1.375 | 0.625 | 0.625 | 0.625 | 0.625 |
| | 2 | 97±7 | 120±8 | 78.0±2.9 | 88.4±2.7 | TI | TI | 1.5 | 1.5 | 0.875 | 0.875 | 1.5 | 1.5 | 0.625 | 0.625 | 0.625 | 0.625 |
| | 3 | 89±9 | 104±9 | 63.2±3.4 | 70.0±3.2 | TI | TI | 1.125 | 1.125 | 0.25 | 0.25 | 1.125 | 1.125 | 0.875 | 0.875 | 0.875 | 0.875 |
| | 4 | 87±7 | 106±8 | 70.0±3.8 | 80.2±3.9 | FE | TI | 1.25 | 1.375 | 0.875 | 0.875 | 1.25 | 1.375 | 0.375 | 0.5 | 0.375 | 0.5 |
| | 5 | 119±7 | 127±7 | 82.6±2.4 | 86.5±2.5 | FE | FE | 1.375 | 1.375 | 0.75 | 0.625 | 1.375 | 1.375 | 0.625 | 0.75 | 0.625 | 0.75 |
| | 6 | 168±7 | 176±9 | 93.8±3.1 | 95.6±1.5 | FE | FE | 1.25 | 0.875 | 0.75 | 0.5 | 1.25 | 0.875 | 0.5 | 0.375 | 0.5 | 0.375 |
| | 7 | 165±9 | 187±7 | 88.0±1.6 | 92.3±1.9 | TI | TI | 1.0 | 1.0 | 0.375 | 0.375 | 1.0 | 1.0 | 0.625 | 0.625 | 0.625 | 0.625 |

| All Classes | Without no control state |
|---|---|

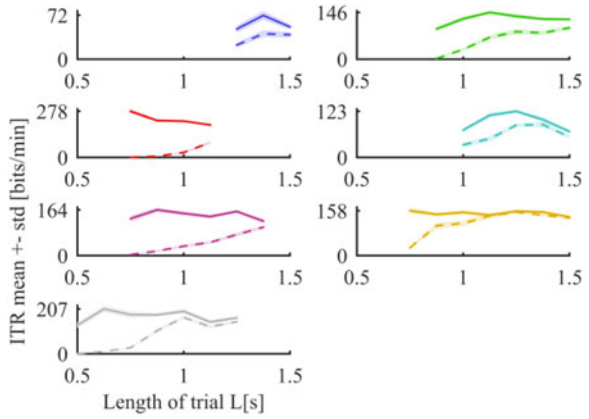


Fig. 12. Comparison of FI (solid lines) and FF (dashed lines) mode ITRs over trial lengths $L$ for each subject (subplots). All 17 classes, TI classifiers, and optimal $l$ and $\Delta t$ parameter values were used for each data point. Ordinates are labeled with the respective maxima.

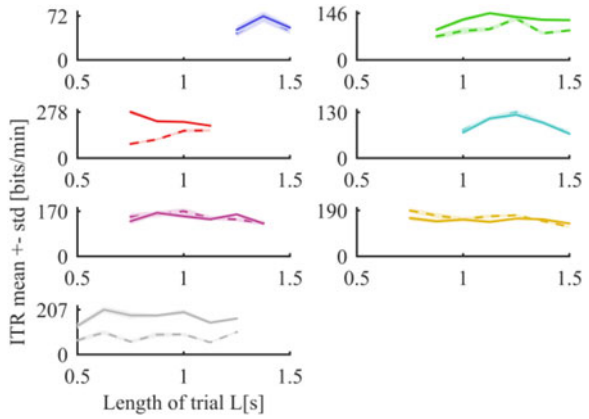


Fig. 13. ITR comparison of TI (solid line) and FE (dashed line) classifiers over trial length $L$ for each subject (subplots). All 17 classes, FI mode, and optimal $\Delta t$ and $l$ parameters were used for each data point. Ordinates are labeled with the respective maxima.

allowed for a broad range of native and accurately displayable stimulation frequencies. In comparison, PC screen-based stimulators feature a broader availability, decreased resolution in time, and usually lack the option to be synchronized with an external clock signal, as well as a highly constant frame rate. Differences in the independent clock sources of the EEG Amplifier and the PC screen are expected to result in inaccuracies of the stimulated and detected frequencies, and should result in decreased feature contrasts.

Our data confirmed ([8]) differences in feature value contrasts between the individual stimulation frequencies. To account for this parameter, we adjusted the stimulation frequencies for each subject by SMCV to yield maximal contrast between target classes. Bin *et al.* [14], as well as recent CCA-based SSVEP BCI implementations ([16], [26], [27]), each applied a maximum criterion for classification of the feature values. However, the authors in [16], [26], and [27] utilized training data, while [14] did not. If training data are available, we propose to account for the differences in baseline power, attended power, and feature contrast ([8]; compare Table I), not only during initial stimulus optimization, but in the classification process as well (here implemented as DQDA).

Improvements targeting components of the CCA method show great potential in recently published SSVEP BCIs. Chen *et al.* [15] divided the ideal sinusoidal reference signals into a filter bank of overlapping subbands in order to access the independent information contained in the harmonics. Zhang *et al.* ([26] and [27]) replaced the artificial reference signals with training data. Our implementation optimized the number (1 or 2) of ideal sinusoidal harmonics in the reference signal. Higher order harmonics contains the potential for further ITR optimization, as pointed out by the reviewers.

We analyzed dynamically stable properties of a continuous SSVEP and trained FE classifiers for data of the FE state using different length of the window of analysis. Our data suggest that the degradation of CCA performance on short data windows

(also reported by Wu [13]) is mainly caused by the spatial filter component of the CCA. Under the assumption that the spatial origin of the SSVEP signal is stationary, this challenge can be overcome by precomputing the spatial filter from long segments of transition-less, FE SSVEP data and then applying it on shorter windows.

Nakanishi *et al.* [16] also optimized the spatial filter component of the CCA ($\sim +70\%$ ITR improvement vs. CCA), but used averaged training data as reference signals, and projected on a bank of spatial filters determined pairwise from all possible combinations of reference signal, test data, and ideal sinusoids. Compared to the method employed in [16], CCAFSF features a decreased complexity of the algorithm (serialized execution of two standard CCAs; single spatial filter and correlation instead of a pairwise filter bank; ideal sinusoidal reference instead of a data driven approach), but comparable, if not increased, efficiency ($+107\%$ VITR improvement versus CCA).

Due to the serialized execution of two standard CCAs in CCAFSF, the method allows for combination with other CCA optimizations targeting different parameter dimensions (e.g., FBCCA [15]). In a similar way, the phase feature extraction implemented here exploited the increased signal contrast in the subspace defined by CCAFSF.

Chen *et al.* [17] recently adopted the method in [16] and incorporated the highly efficient (267 Bits/min) JFPM stimulus encoding technique.

Comparison of the currently fastest published SSVEP BCIs ([15]–[17], [26], [27]) reveals a number of evolving functional components (stimulus encoding, reference signal design, spatial filter training, classification) recently centered around improvements to the standard CCA algorithm [14]. This development suggests a not yet fully exploited optimization potential of the individual components and even more so their optimal combination.

Our analysis showed that the implementation specific upper limit of the VITR, resulting from classification of data in the FE state, can reach up to $\sim 1\mathrm{kBit}/\mathrm{min}$ in some subjects. Such rates will not be achieved in practical online systems, though, because the switching between targets causes transients in the SSVEP response.

We analyzed this transient activity using a dedicated protocol. A visual cueing method indicated the next target to attend to, once the next short auditory beep was perceived. The auditory cue was given in constant intervals and classification result neutral (beep) similar to a metronome. This method avoids searching the target layout for the next item to be entered by simply marking it by color. In this manner, faulty memory recalls and delays associated with searching or recalling the next target are avoided and the duration of continuous input is extended. We consider this a model for a real life application by a subject who is firmly trained on the device and who produces input on own intend while still randomizing the target sequence.

The transients were shown to consist of dead times, the SSVEP entrainment processes, and delays, which each constituted a negative impact to the system‘s ITR. To lessen this impact, we optimized the time available for each input $L$, the length of the window of analysis $l$, its offset $\Delta t$ relative to the auditory cue, and the classifier training set (FE or TI). Furthermore, we compared the application of two different modes of operation (FI and FF) and two sets of targets (with and without the no-control state). The ITR optimization was implemented as an exhaustive search and left system accuracy as a driven, dependent value. In BCI applications in which a higher than ITR effective accuracy is desirable, an additional accuracy constrained can be added to the exhaustive search.

A comparison of classifier performance when using training data taken from the FE SSVEP (FE classifier) versus training data taken from the interval when the SSVEP get entrained (TI classifier) revealed that if the TI classifiers outperform the FE classifiers for the individual subject, the resulting ITR being among the highest of the population tested. We hypothesize that the subject‘s ability to produce highly constant reaction times to the auditory cue is a requirement for efficient TI classifier operation, and that this behavior can be trained.

We evaluated the system’s performance with and without using the no control state class and found that detection of the no control state has a negative impact on the ITR; thus, it should only be used if required by the application.

A key innovation in our study, which is not restricted to the analysis or paradigm used here, is the introduction of an FI operation mode for BCI. The delays of a closed-loop system add to the time required per roundtrip. In a BCI where feedback of the classification result cues for the next input (FF mode), these delays directly add to the time required per input but do not contribute information to the classification process. This resembles the inefficiency incurred from visually validating the result of every keystroke before performing the next when typing on a keyboard.

By separating the system-to-subject communication into two components—feedback of the classification result (data) and input cue signal (clock), we effectively decoupled the subjects BCI input stream from the classification result feedback stream. In the FI mode, the now independent cue signal was issued early and preceded classification to increase the system‘s input bandwidth by shifting the delays from the BCI‘s input (target selection) to its output (classification result feedback). The resulting ITR increase comes at the cost of the disability to react to the very last classification result in the next input. The ability to react to feedback is only delayed, however, so the possibility to react to the second to last or older classification results (depending on the fraction of time required per input and the delays) remains. The reaction to delayed feedback requires covert attention to the feedback stream as well as synchronization of the next (correcting) input with the next cue. We consider the visual cueing method (red LED) implemented here as a model for this simultaneous covert attention task.

Depending on the BCIs eventual application, the FI or FF mode remain a choice of design. Sometimes, it might be preferable to perform the next input as soon as ITR effective at the cost of delayed feedback (FI mode). Other times, it can be preferable to idle the classification process and drop data until classification feedback was perceived, reacted to, and resulted in the next SSVEP state change, in order to enable seemingly “immediate”

(from the perspective of the subject) reaction to feedback at the cost of reduced ITRs (FF mode).

In this study, feedback was not given during any of the experiments. The subjects instead reacted to an audio cue that did not carry information about the classification result. Feedback about a false classification result is expected to introduce additional reaction time delays. The FI and FF ITR results presented here do not take these additional delays into account, but only measured the unidirectional subject to system information transfer at the output of the Level 1 BCI control module [24].

Cue-guided spelling allows both to clock the BCI's input at a predetermined optimal ITR rate and to circumvent the negative effect of the system's delays on the input bandwidth if issued early (FI mode). Aligning the window of analysis relative to the input cue is assumed to enable classification with increased precision, as the time at which the input occurs does not have to be deducted from the data. The same cue-guided input restrictions which benefit the receiver, however, hinder the sender, limit the possible input correction strategies, and constitute an overall less intuitive, more difficult to perform, input scheme.

The ratio of the optimal length of the window of analysis and the time available for each input revealed that only $\sim 49.8\ \%$ of the total exposition to stimuli was informative for classification in the FF mode. This result matches the bandwidth utilization reported in the currently fastest SSVEP BCI implementation [17] (0.5 s stimulation followed by 0.5 s target selection).

The analysis of the FI mode data revealed an improvement of this ratio to $\sim 22.5\%$ by minimizing the negative impact of the system's delays on the input bandwidth. The remaining losses represent an effective dead time of the detectable SSVEP response (less informative than efficient for classification) and are expected to define the system-specific maximum input rate. Methods to further reduce these bandwidth losses appear highly desirable for further ITR improvements.

The main difference between the FF and the FI mode resulted from the delays of the system and subject. Using a low latency hard- and software implementation allows to nearly eliminate the system's contribution to this discrepancy. However, the latency introduced by the subject ($t_{react}$) comprised of the reaction time and latency of the visual pathway ($\sim 140$ ms according to [17]) remains. In the FI mode, these remaining latencies can be circumvented at the cost of delayed feedback.

We demonstrated the considerable impact of the delays on FF mode ITRs. These delays, however, are usually not reported in BCI studies ([15], [16], [19]). The reported ITRs are consequently compromised by considerable uncertainty. FI mode ITR results are not biased by the individual system delays and thus easier to interpret and considered to be better suited to compare the performance of presented algorithms.

The reference signal is used in the CCA models SSVEPs that have constant amplitude over the whole window of analysis. We expect that consideration of the dynamics of the SSVEP power envelope during transition periods would allow access to additional information from the interval appearing as effective dead time to the constant amplitude model. Such a method, however, is yet to be developed.

## VI. Conclusion

Our component-wise analysis of an FDMA SSVEP BCI reveals an upper limit for the ITR at 490 Bit/min on average, reaching up to 1 kBit/min in individual subjects. We analyze why current FDMA SSVEP BCIs and our FF mode implementation (average 124 Bit/min; peak $187 \pm 7\,\mathrm{Bit/min}$) fall short of these rates, and present methods that address them. The resulting FI BCI system increases the bandwidth available for subject-to-system transmission and thus ITR (average 181 Bit/min with $295 \pm 12\,\mathrm{Bit/min}$ peak performance) by circumventing input latencies of the BCI system at the cost of delayed classification result feedback. We introduce the CCAFSF feature extraction method and demonstrate a $+107\%$ VITR increase on data of the FE SSVEP when compared to the standard CCA [14]. Our analysis of the transitions between targets reveals a dead time of the SSVEP response which defines a maximum input rate for the BCI and is shown to be further responsible for the drop in ITR from the upper limit, when the SSVEP is FE, to the online system. We therefore suggest to develop feature extraction methods which specifically evaluate the transitions between SSVEP states, rather than considering the SSVEP as constant in power over the whole window of analysis.

## References


[1] D. Zhang, A. Maye *et al.*, "An independent brain-computer interface using covert non spatial visual selective attention," *J. Neural Eng.*, vol. 7, no. 1, pp. 1436–1447, Feb. 2010.

[2] S. Gao, Y. Wang *et al.*, "Visual and auditory brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1436–1447, May 2014.

[3] R. Singla, A. Khosla *et al.*, "Influence of stimuli color on steady-state visual evoked potentials based BCI wheelchair control," *J. Biomed. Sci. Eng.*, vol. 6, no. 11, pp. 1050–1055, Nov. 2013.

[4] V. Odom *et al.*, "Visual evoked potentials standard (2004)," *Doc. Ophthalmol.*, vol. 108, no. 2, pp. 115–123, Mar. 2004.

[5] P.-L. Lee, C.L. Yeh *et al.*, "An SSVEP-based BCI using high duty-cycle visual flicker," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3350–3359, Dec. 2011.

[6] F. Guo, B. Hong *et al.*, "A brain-computer interface using motion-onset visual evoked potential," *J. Neural Eng.*, vol. 5, no. 4, pp. 477–485, Dec. 2008.

[7] S. Schaeff, M. S. Treder *et al.*, "Exploring motion VEPs for gaze-independent communication," *J. Neural Eng.*, vol. 9, no. 4, pp. 045006, Aug. 2012.

[8] L. Piccini, S. Parini *et al.*, "A wearable home BCI system: Preliminary results with SSVEP protocol," in *Proc. IEEE 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2005, pp. 5384–5387.

[9] G. Bin, X. Gao *et al.*, "A high-speed BCI based on code modulation VEP," *J. Neural Eng.*, vol. 8, no. 2, pp. 025015, Apr. 2011.

[10] G. Bin, X. Gao *et al.*, "VEP-based brain-computer interfaces: Time, frequency, and code modulations," *IEEE Comput. Intell. Mag.*, vol. 4, no. 4, pp. 22–26, Nov. 2009.

[11] H. Wang, T. Li *et al.*, "Remote control of an electrical car with SSVEP-based BCI," in *Proc. IEEE Int. Conf. Inf. Theory Inf. Security*, 2010, pp. 837–840.

[12] S. Parini, L. Maggi *et al.*, "A robust and self-paced BCI system based on a four class SSVEP paradigm: Algorithms and protocols for a high-transfer-rate direct brain communication," *Comput. Intell. Neurosci.*, vol. 2009, Jan. 2009, Art. no. 864564, 11 pages.

[13] Z. Wu, "SSVEP extraction based on the similarity of background EEG," *PLoS One*, vol. 9, no. 4, pp. e93884, Apr. 2014.

[14] G. Bin, X. Gao *et al.*, "An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method," *J. Neural Eng.*, vol. 6, no. 4, pp. 046002, Aug. 2009.

[15] X. Chen, Y. Wang *et al.*, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface," *J. Neural Eng.*, vol. 12, no. 4, pp. 046008, Jun. 2015.

[16] M. Nakanishi, Y. Wang *et al.*, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 6, pp. 1450019, Sep. 2014.
[17] X. Chen, Y. Wang *et al.*, "High speed spelling with a non-invasive brain-computer interface," *PNAS*, vol. 112, no. 44, pp. E6058–67, Sep. 2015.
[18] G. Volosyak, "SSVEP-based Bremen-BCI interface-boosting information transfer rates," *J. Neural Eng.*, vol. 8, no. 3, pp. 036020, Jun. 2011.
[19] M. Spüler, W. Rosenstiel *et al.*, "Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning," *PLoS One*, vol. 7, no. 12, pp. e51077, Dec. 2012.
[20] P. Yuan, X. Gao *et al.*, "A study of the existing problems of estimating the information transfer rate in online brain-computer interfaces," *J. Neural Eng.*, vol. 10, no. 2, pp. 026014, Apr. 2013.
[21] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL, USA: Univ. Illinois Press, 1964.
[22] C. Jin, X. Gao *et al.*, "Frequency and phase mixed coding in SSVEP-based brain-computer-interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 200–206, Jan. 2011.
[23] D. Regan, "Some characteristics of average steady-state and transient responses evoked by modulated light," *Electroencephalogr. Clin. Neurophysiol.*, vol. 20, no. 3, pp. 238–248, Mar. 1966
[24] D. E. Thompson, S. Blain-Moraes, J. E. Huggins, "Performance assessment in brain-computer interface-based augmentative and alternative communication," *Biomed. Eng. Online*, vol. 12, no. 1, pp. 1, May 2013.
[25] X.D. Zhang, "*Optimal High-Throughput Screening: Practical Experimental Design and Data Analysis for Genome-scale RNAi Research*," New York, NY, USA: Cambridge Univ. Press, 2011.
[26] Y. Zhang, G. Zhou *et al.*, "L1-regularized multiway canonical correlation analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 887–896, Nov. 2013.
[27] Y. Zhang, G. Zhou *et al.*, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 4, pp. 1450013, Jun. 2014.

Authors' photographs and biographies not available at the time of publication.