

## How fast can f-VEP BCIs ever be?

Malte Sengelmann, Andreas K. Engel, and Alexander Maye

**Abstract**— VEP-based BCIs offer a number of advantages that make them a promising candidate for applications in everyday environments: They do not require user training, the high signal-to-noise ratio of VEPs allows reliable classification, and high information transfer rates (ITRs) of up to ~100 bits/min have been achieved during recent years. In this article we estimate an upper bound of the ITR for VEP BCIs that use frequency and phase coding for classification (f-VEP BCIs). The estimate is based on an idealized classification process that operates on real EEG data of the steady-state (SSVEP) from naïve subjects. Our study yields subject-specific upper bounds in the range of approx. 200 to 500 bits/min. We identify causes for the significantly lower ITR of existing f-VEP BCIs and suggest solutions that can narrow the gap to the upper bound.

### I. INTRODUCTION

One of the most investigated paradigms for Brain-Computer Interfaces (BCIs) exploits the electrical potentials that are evoked by low-frequency flickering light (Visual Evoked Potentials, VEPs) for classifying user intent. The user selects a target by directing attention overtly, i.e. by gazing, or covertly [1] to a flickering light source. This attentional shift increases the ratio between the signal that is encoded in the light source and the noise of the background brain activity, which in turn allows to recognize the selected target by a classifier. VEP-based BCIs differ in the way how channels are encoded [2, 3]. Here we consider a system where the channel identity is encoded in the frequency and phase of a set of steadily flickering LEDs (f-VEP).

Several approaches to improve the usability and reliability of VEP-based BCIs have been developed in recent years. The visual fatigue caused by high-intensity flicker has been countered by using stimuli that invert their contrast [4] or switch between stationary and moving patterns [5, 6]. Classification accuracy has been increased by using subject-specific stimulation frequencies [7] and various filter techniques like common spatial patterns (CSP) [8] or canonical correlation analysis (CCA) [9].

Classification latency is another parameter with a significant impact on the usability of a BCI, but methods for increasing the Information Transfer Rate (ITR) of VEP-based BCIs have been less studied. The currently fastest BCIs achieve average ITRs between 92.8 and 124 Bits/min [2, 3, 10].

We were interested in the question how the latency of an f-VEP BCI could be minimized and hence its ITR be maximized. To answer this question we analyzed which

processes affect the measurement of the ITR. This analysis considered the full setup consisting of the user, the task, the hardware, and the methods for signal processing and classification. Next we separately analyzed the latencies that inherently determine the ITR and latencies caused by the setup and the subject. We then collected experimental data and determined offline the dependency of the ITR on the stimulation frequency and sample size used for classification. As expected we observed a trade-off between classification accuracy and sample size. We found significant performance differences between users as well as between reported online ITRs [2, 3, 10] and our estimate for an upper bound of the ITR. Whereas we can only speculate about the causes of the inter-individual performance differences, we discuss the grounds for the gap between online and offline ITRs.

### II. LATENCIES IN AN F-VEP BCI

We begin by identifying the processes that affect ITR measurements in f-VEP BCIs. In a typical setup the subject is asked to select a number of targets in a given order. This order is given just prior to response, e.g., by asking to enter a certain number or word, or in advance, by announcing the next target when classification of the previous has finished. In both cases the subject needs some time to prepare (e.g., think or find out which target to select next) and execute (e.g., move the eyes) the corresponding action. We call this a reaction time  $t_{react}$ . Once the gaze is focused on a specific target, it takes some time to entrain the cortical network so that an f-VEP can be detected ( $t_{entrain}$ ). Most current EEG amplifiers are not optimized for real time operation. There are numerous queues, buffers, interrupts and processing times of the equipment involved before the data reaches the analyzing algorithm ( $t_{equip}$ ). Next a number of data samples have to be collected for a reliable estimate of the power at the respective frequency ( $t_{sample}$ ). Finally the classification process and the generation of user feedback requires the time  $t_{classify}$  before the next iteration of the whole cycle begins. The time needed to select one target in the BCI hence is the sum of these delays,

$$t_{select} = t_{react} + t_{entrain} + t_{equip} + t_{sample} + t_{classify}, \quad (1)$$

and the ITR is determined by the number of bits transferred per selection divided by this time.

When maximizing the ITR of an f-VEP BCI, only  $t_{sample}$  and  $t_{classify}$  are under direct control of the BCI engineer. Reaction times can be minimized by using salient cues, but they inherently depend on the characteristics of the subject. This is particularly true for  $t_{entrain}$ . Likewise  $t_{equip}$  can be minimized by choosing appropriate hardware, but it is fixed for a given setup. We will deal with  $t_{classify}$  in a forthcoming publication and focus here on the optimization of  $t_{sample}$ .

\* This study has been supported by grants from the EU (FP7-ICT-270212, ERC-2010-AdG-269716) and the DFG (GRK 1247/1/2). The authors are with the Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany (e-mail: {m.sengelmann, ak.engel, a.maye}@uke.de).

When continuously estimating power over time, the resolution in time and frequency are interdependent: Longer sample windows increase the accuracy of power estimation and hence classification at the expense of classification rate. In the following sections, we analyze this tradeoff and determine the subject-specific sample sizes  $t_{\text{sample}}$  that maximize the ITR, neglecting all remaining delays.

### III. METHODS

#### A. Hardware

Visual stimulation was generated by an array of 16 LEDs, each with a conical diffusor lens of 3 cm diameter (visual angle of 5°). Elements were arranged 4x4 with a grid distance of 6 cm (10° visual angle). At the left edge of the LED array there was a column of 4 static fixation spots for the no control state. They were placed on the same grid as the LEDs, extending the matrix to 5x4. The LEDs were driven by rectangular signals with a duty cycle of 50%. These signals were recorded as one trigger signal per LED by the EEG amplifier.

#### B. EEG recording

EEG signals from the 14 electrodes that are closest to position Oz in the 10-20 system were recorded with a BioSemi ActiveTwo amplifier at 1024 Hz sampling frequency. Seven healthy subjects in the age between 26 and 44 (mean 29 years) were seated upright in a relaxed position on a chair in front of the LED array (distance 35 cm). The recording took place in a dimmed environment with some residual light coming from a static image of a 15" TFT screen placed outside the field of view of the subject.

#### C. Experiment conditions and task

In the 1<sup>st</sup> session, subject-specific stimulation parameters were determined. Subjects were asked to overtly attend to each of the simultaneously presented targets on the matrix in a given sequence. Data from 50 s of continuous exposure to each stimulus without shifting gaze or attending to another stimulus were recorded in order to capture the SSVEP. After an optional period of relaxation, subjects proceeded to the next target until each of them was attended once. Each target was driven at 1 of 16 different frequencies ranging from 11 Hz to 43 Hz in 2 Hz steps skipping 25 Hz. This distribution keeps a band of  $\pm 1$  Hz around every 1<sup>st</sup> and 2<sup>nd</sup> harmonic of stimulus frequencies which is free from other stimulus frequencies and line noise.

In the 2<sup>nd</sup> session, the 16 LEDs plus the 4 fixation spots were used as the 20 targets. The targets were processed in the same way as described for the 1<sup>st</sup> session. The LEDs were driven with 4 different frequencies in 4 relative phase angles of 0°, 90°, 180° and 270° each. The frequencies were selected to maximize the ITR of the respective subject based on results of the 1<sup>st</sup> session.

#### D. Signal processing

EEG data were processed with a finite impulse response (FIR) minimum order equiripple band pass filter with -80dB attenuation in the stopbands and corner frequencies at the lowest stimulus frequency minus 2 Hz and the highest stimulus frequency times 2 plus 2 Hz. This filter setting retains the 1<sup>st</sup> and 2<sup>nd</sup> harmonics of every stimulation

frequency in the passband. The filter is restarted at the beginning of each recording and its startup time is removed from further processing. The data are cut in windows of a fixed length L with an overlap of L-1/8 s. Two CCAs, one using only the 1<sup>st</sup> harmonic and another using the first two harmonics, were applied to data in each window and yielded two canonical correlations R as two scalar features [9]. The complex Fourier coefficients of the stimulus frequencies were extracted with a Fast Fourier Transform (FFT) from channel Oz. Subsequently their real and imaginary parts were used as features for phase detection. Triggers of the four LEDs running at 0° phase were used as reference.

#### E. Stimulus and feature extraction method performance measure

The 1<sup>st</sup> session was used to determine for each subject the optimal stimulation frequencies [7] and one corresponding feature extraction method for each of them that maximizes group difference [11]. The method used here is the Standardized Mean of a Contrast Variable (SMCV) [12]. Size of effect values  $\lambda$  were computed for each of the 16 stimulus frequencies and both CCA types with a window length of L=2 s. We contrasted the feature values with and without attention to the respective stimulus. A larger  $\lambda$  indicates a feature in which the mean values of the two contrasted groups differ more and/or the standard deviations of the two groups are smaller. The 4 frequencies and corresponding feature extraction methods resulting in the largest among the 32 tested  $\lambda$  values were selected for the 2<sup>nd</sup> session. These parameters were considered as optimal in terms of classification accuracy.

#### F. Classifier

Classification was performed in two steps: The first classifier distinguished between the control state and the no control state using one scalar feature and feature extraction method per frequency. If a control state was detected, the phase was classified using the phase features of the detected frequency in the second step. Both steps used a linear discriminant analysis (LDA).

#### G. ITR performance measure

The 2<sup>nd</sup> session data were re-analyzed 20 times for different settings of the window length parameter L ranging from 1/8 s to 20/8 s in steps of 1/8 s. The data used in each of the 20 steps are called a dataset. The number of sample windows in each dataset is held constant. For each L, the corresponding dataset is separated in three subsets for training, validation and optimization. Classifiers were trained on the training subset. The optimization subset was used to adapt a spatial filter. The validation subset is used to report results.

Performance was measured by ITR in bits per minute for each of the 20 datasets as follows [13, 14]:

$$B = \text{ld } N + P \text{ld } P + (1 - P) \text{ld } \left(\frac{1-P}{N-1}\right) \quad (2)$$

$$\text{ITR} = B * (60/L) \quad (3)$$

B is the bit rate in bits per symbol, N is the number of classes, and P is the classification accuracy.

### H. Spatial filter

To train the spatial filter, we started with 14 channels and iteratively removed one at a time until the performance stopped to increase. Using the dataset with the maximum ITR on the optimization subset, the next channel to be tested for removal was selected. The selection was based on the canonical coefficients  $\mathbf{A}$  of the EEG channels calculated during CCA feature extraction.  $\mathbf{A}$  has the dimension  $n$  by  $s$  by  $i$ , where  $s$  is the number of sinusoids used in the CCA template ( $s = 2$  for a 1<sup>st</sup> harmonic template and  $s = 4$  for a 1<sup>st</sup> and 2<sup>nd</sup> harmonic template),  $n$  is the current number of channels, and  $i$  is the number of data windows. For each data window,  $\mathbf{A}$  was computed using only the CCA template that matches the corresponding stimulation frequency. The weights  $W_{n,i}$  were calculated from  $\mathbf{A}$  for each  $n$  and  $i$  as the Euclidean norm over the  $s$  dimension.  $W$  is then averaged over all trials of the current dataset.

$$W_{n,i} = \sqrt{A_{1,n,i}^2 + A_{2,n,i}^2 + \dots + A_{s,n,i}^2} \quad (4)$$

$$w_n = (W_{n,1} + W_{n,2} + \dots + W_{n,i})/i \quad (5)$$

The channel with the minimum in  $w_n$  is considered to contribute the least information and is removed from every dataset resulting in 20 new datasets to be used in the next iteration step. If the ITR maximum of the current step was lower than the ITR maximum of the last, the iteration was aborted. The resulting channel set with the maximum overall ITR constituted the spatial filter.

### IV. RESULTS

The SMCV values  $\lambda$  for each stimulation frequency and feature extraction method resulting from the 1<sup>st</sup> session are plotted in Fig. 1 for a representative subject. The figure shows that the attentional modulation of f-VEP power is strongest at 19, 21, 29 and 31Hz, and these frequencies are selected as the optimal stimulation parameters for this subject.

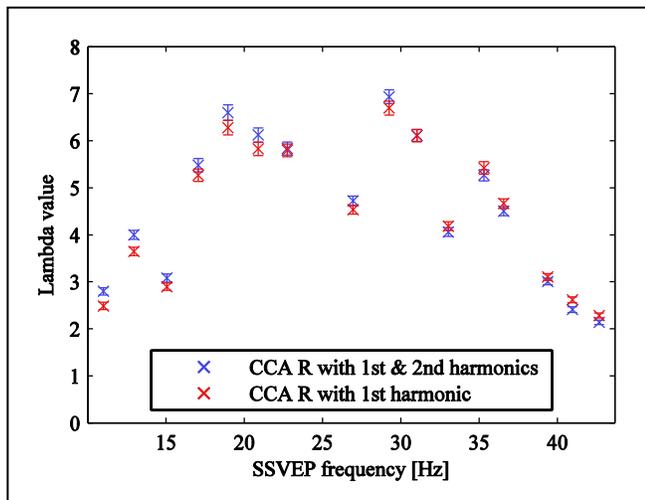


Figure 1. SMCV  $\lambda$  values over frequency for subject 3. Bootstrapping (omitting 25% of the data,  $n=1000$ ) was used to estimate the variance. All differences are significant at  $p < 0.01$  except for 31Hz.

Fig. 2 shows the ITRs as a function of window size  $L$  for the same subject calculated on the 2<sup>nd</sup> session data with and

without spatial filtering. The ITRs of the training, validation and optimization subset show only negligible differences. This indicates that the classifier and the spatial filter were not over-fitted. Spatial filtering increases the ITR until the classification accuracy approaches 100% for window lengths over 700ms.

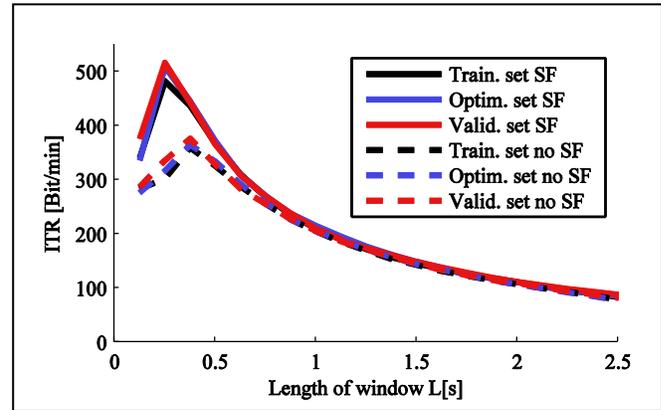


Figure 2. ITR as a function of window size  $L$  for subject 3 with and without SF on training, validation and optimization subsets

Fig. 3 shows the ITRs for all subjects based on the 2<sup>nd</sup> session validation datasets. The window sizes where the ITR peaks are subject-specific and lie in the range from 250-625ms.

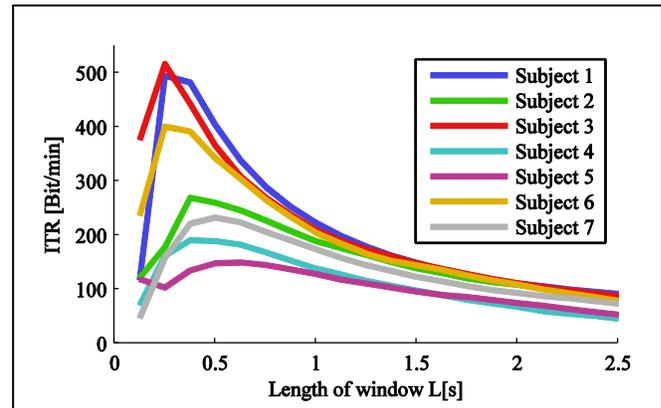


Figure 3. ITR as a function of window size  $L$  on the validation subsets with spatial filtering for all subjects

Table 1 summarizes key values for all subjects. The magnitudes of the four largest  $\lambda$  values for each subject describe the general ability to modulate the SSVEP at the respective stimulus frequency by attention. Comparing the average  $\lambda$  values with the maximum ITR shows that larger  $\lambda$  values tend to result in higher ITRs. However, there is no direct relation, because  $\lambda$  values are computed on 1<sup>st</sup> session data whereas the ITR is calculated on 2<sup>nd</sup> session data.

TABLE I. SUBJECT PERFORMANCE OVERVIEW

Subject Nr.	Performance											
	Best stimulation frequencies $f$ [Hz] and corresponding $\lambda$ value								Average of best four $\lambda$ s	Maximum ITR [Bits/min]	$L$ at maximum ITR [ms]	ITR increase by spatial filtering
	$f1$	$\lambda1$	$f2$	$\lambda2$	$f3$	$\lambda3$	$f4$	$\lambda4$				
1	13	8.8	15	7.1	17	9.0	21	9.2	8.5	495	250	19.7%
6	17	7.2	19	7.6	23	8.0	31	7.7	7.6	401	250	13.1%
2	11	6.4	13	6.9	15	8.0	17	6.2	6.9	269	375	12.0%
3	19	6.6	21	6.1	29	7.0	31	6.2	6.5	517	250	37.2%
7	15	4.1	17	5.4	19	4.4	21	4.8	4.7	233	500	20.4%
5	17	3.9	19	4.4	21	4.4	31	3.3	4.0	150	625	3.8%
4	11	2.9	17	3.1	19	3.2	35	3.1	3.0	191	375	15.7%

## V. DISCUSSION

This study identifies three parameters with a major impact on the ITR of an f-VEP BCI: the stimulation frequency, the size of the analysis window, and the spatial filter. For each parameter we propose an optimization method.

For stimulation, frequencies should be selected that are maximally modulated by the attention of the user. We use the group difference index SMCV to compare the feature distributions of SSVEP power for the two conditions “stimulus attended” and “baseline”. Employing this method at the level of feature values, optimal stimulation frequencies and feature extraction methods can be selected without depending on the type and properties of specific classifiers.

To optimize the size of the analysis window, we use data from the steady-state and analyze the dependency of the ITR from this parameter. In comparison with the window lengths that are typically used for f-VEP BCIs we find that the maximum ITR for the individual subjects is reached for surprisingly short windows. The size of the analysis window critically determines the upper bound for the ITR. Taking a closer look at (3) one notices that adding a fixed cost or latency as discussed in section II. will only change the scale in the plots in Figs. 2 and 3, leaving the subject specific ITR peak in the relative same location. This underlines the relevance of the ITR maxima found here for some cases of  $t_{\text{select}}$  which could not be discussed in the scope of this paper.

We optimize a spatial filter by iteratively removing channels with low information until increases in ITR level off. Removing channels in general is inferior to assigning small weights as it reduces the total amount of information in the data. In our approach the CCA method itself already constitutes a spatial filter. It is therefore ineffective to compute another channel weighting unless these weights are either one or zero. The selection of channels is based on the information of the full dataset. The binary spatial filter we present here reduces the noise in channel weights that is incurred by applying the CCA on short windows of length  $L$ . This results in an increased ITR.

## VI. CONCLUSION

We suggest that the maximum ITR in the steady-state, i.e. when the VEP is fully entrained, marks an upper bound for any f-VEP BCI. This ITR depends on the classification accuracy and thus on the signal to noise ratio (SNR) of the features which can still be improved using different methods. The maximum steady-state ITRs determined in this study

reach ~200-500 bits/min, which indicates a high potential of this BCI paradigm that is not yet fully exploited.

A large part of the gap between this upper bound and the performance of current f-VEP BCIs derives in our view from the fact that these systems have been mainly optimized for the steady-state without addressing transitions between those states, i.e. without considering  $t_{\text{entrain}}$  in our nomenclature. The CCA method uses a sinusoid which is “steady” in amplitude as a template to correlate with the data. The high ITRs we find in this study suggest that these templates match the steady-state data well. In contrast, data from online BCIs contain additional epochs of transition between targets. The properties of these epochs and their relation to the steady-state have to be analyzed in order to be able to reliably recognize steady-states and apply the classification on their data. We propose to divide the challenge of high-ITR BCIs into two sub-problems. First, data without transitions should be used to test feature extraction methods and gather subject specific parameters for each target state. Second, the transitions between steady-states should be detected and used to align the classification window to epochs of steady-states. This allows the CCA template and the classifiers, which were trained on the steady-state, to match the live data well and thereby to maximize classification accuracy. Results of our work that exploit temporal properties of the entrainment of SSVEPs for online BCIs are forthcoming.

## ACKNOWLEDGMENT

We would like to thank Dan Zhang of Tsinghua University, Beijing, for his valuable comments on this study.

## REFERENCES

- [1] D. Zhang, A. Maye, X. Gao, B. Hong, A.K. Engel and S. Gao, *An independent brain-computer interface using covert non spatial visual selective attention*, J. Neural Eng. 7, 2010
- [2] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao, *A high-speed BCI based on code modulation VEP*, J. Neural Eng. 8, 2011.
- [3] G. Bin, X. Gao, Y. Wang, B. Hong, and S. Gao, *VEP-Based Brain-Computer Interfaces: Time, Frequency, and Code Modulations*, IEEE Computational Intelligence Magazine, 2009
- [4] V. Odom, M. Bach, C. Barber, et al., *Visual evoked potentials standard (2004)*, Documenta Ophthalmologica, 108, no. 2, pp. 115–123, 2004.
- [5] F. Guo, B. Hong, X. Gao, and S. Gao, *A brain-computer interface using motion-onset visual evoked potential*, J. Neural Eng. 5, 2008
- [6] S. Schaeff, M. S. Treder, B. Venthur and B. Blankertz, *Exploring motion VEPs for gaze-independent communication*, J. Neural Eng. 9, 2012
- [7] L. Piccini, S. Parini, L. Maggi and G. Andreoni, *A Wearable Home BCI system: preliminary results with SSVEP protocol*, Proceedings of the 2005 IEEE Eng. in Medicine and Biology 27<sup>th</sup> Conference, 2005

- [8] S. Parini, L. Maggi, A. C. Turconi and G. Andreon, *A Robust and Self-Paced BCI System Based on a Four Class SSVEP Paradigm: Algorithms and Protocols for a High-Transfer-Rate Direct Brain Communication*, Computational Intelligence and Neuroscience 9, 2009, Article ID 864564
- [9] G. Bin, X. Gao, Z. Yan, B. Hong and S. Gao, *An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method*, J. Neural Eng. 6, 2009
- [10] G. Volosyak, *SSVEP-based Bremen-BCI interface-boosting information transfer rates*, J. Neural Eng. 8, 20110
- [11] C.J. Huberty, *A history of effect size indices*, Educational and Psychological Measurement 62: 227–40, 2002, 2002
- [12] J. Cohen, *The statistical power of abnormal-social psychological research: A review*, Journal of Abnormal and Social Psychology 65: 145–53, PMID 13880271
- [13] J. R. Wolpaw, H. Ramoser, D. J. McFarland and G. Pfurtscheller, *EEG-Based communication: improved accuracy by response verification* IEEE Trans. Rehabil. Eng. 6 326–33, 1998
- [14] C. E. Shannon and W. Weaver, *The mathematical Theory of Communication*, Urbana, IL: University of Illinois Press, 1964