

Assessing Perceived Image Quality Using Steady-State Visual Evoked Potentials and Spatio-Spectral Decomposition

Sebastian Bosse, Laura Acqualagna, Wojciech Samek, *Member, IEEE*, Anne K. Porbadnigk, Gabriel Curio, Benjamin Blankertz, Klaus-Robert Müller, *Member, IEEE*, and Thomas Wiegand, *Fellow, IEEE*

Abstract—Steady-state visual evoked potentials (SSVEP) are neural responses, measurable using electroencephalography (EEG), that are directly linked to sensory processing of visual stimuli. In this study, SSVEP are used to assess the perceived quality of texture images. The EEG-based assessment method is compared to conventional methods and recorded EEG data is correlated to obtained mean opinion scores (MOS). A dimensionality reduction technique for EEG data called spatio-spectral decomposition (SSD) is adapted for the SSVEP framework and used to extract physiologically meaningful and plausible neural components from the EEG recordings. It is shown that the use of SSD not only increases the correlation between neural features and MOS to $r = -0.93$, but also solves the problem of channel selection in EEG-based image quality assessment.

Index Terms—EEG, SSVEP, video quality assessment, classification, MOS, spatio-spectral decomposition

I. INTRODUCTION

DIGITAL images and video have become ubiquitous today. Most images and videos that are captured, transmitted and displayed are intended to be ultimately viewed by humans. Channels in image or video transmission systems are bandwidth limited. This limitation necessitates bit rate reduction of the data achieved by compression. However, bit rate reduction comes at a price and with decreasing bit rate, compression algorithms introduce degradations into the signal that are visible to human viewers. In order to deal with the tradeoff between bit rate permitted by the channel capacity (or other infrastructure limitations) and the quality of the visual signal, the measurement of these degradations in a

perceptually relevant manner is crucial for the operation of those transmission systems.

The assessment of perceived visual quality and the underlying fundamental model of human vision has been an ongoing research area since decades [1]. Although a lot of progress has been made in recent years, a precise model for perception of visual distortion is not available and the question how to quantify perceptual visual quality remains answered unsatisfyingly [1].

Hence, the assessment and quantification of perceived quality and the evaluation of transmission systems or modules of transmission systems, such as compression schemes, is commonly still approached in psychophysical tests. During these tests, a participant (also called: observer) is presented a stimulus in form of an image or a video and gives an overt response on the subjective judgment on the visual quality of the specific stimulus. Test procedures have been standardized for television applications in [2] and for multimedia applications in [3]. The psychophysical assessment of perceived quality may be done based on the rating of a single stimulus via *Absolute Category Rating (ACR)* or based on the comparison of a reference stimulus with a test stimulus via *Degradation Category Rating (DCR)*. The ratings of individual test participants, observers, are pooled by averaging and reported as mean opinion scores (MOS). It is recommended to average over ratings collected from at least 15 observers [2]. Psychophysical quality assessment tests are very exhausting for participants and in order to prevent influence of fatigue, it is further recommended to restrict the duration of a rating session to not more than 30 min [2]. Individual ratings are highly variable across subjects. Also, other factors than the characteristics of the stimuli presented during test sessions have an influence on the individual ratings, since internal states of the participants, such as motivation, fatigue, rating strategies, may have impact on the conscious decision process.

These drawbacks and the desire to gain insight into the internal mental process during decision making in quality assessment tasks motivated researchers to study psychophysiological approaches recently.

This paper presents an image quality assessment study based on an electroencephalographic technique called steady-state visual evoked potentials (SSVEP). The data corpus used for the evaluations presented was used earlier in three other studies addressing different scientific questions [4], [5], [6]. [4] studies the detection of perceived quality degradation in

SB and LA contributed equally.

We acknowledge financial support by the BMBF Grant N. 01GQ0850.

KRM acknowledges partial financial support by BK21 of NRF.

Correspondence should be addressed to SB, BB, KRM, TW.

S. Bosse and W. Samek are with the Video Coding and Analytics department of Fraunhofer Heinrich Hertz Institute (Fraunhofer HHI), 10587 Berlin, Germany (email: sebastian.bosse@hhi.fraunhofer.de).

L. Acqualagna and B. Blankertz are with the Neurotechnology Group, Berlin Institute of Technology, 10587 Berlin, Germany (email: benjamin.blankertz@tu-berlin.de).

Anne K. Porbadnigk is with the Machine Learning Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany.

G. Curio is with the Department of Neurology and Clinical Neurophysiology, Charité, 10117 Berlin, Germany.

K.-R. Müller is with the Machine Learning Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany, and also with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (e-mail: klaus-robot.mueller@tu-berlin.de).

T. Wiegand is with Fraunhofer HHI, and with the Image Communication Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany (e-mail: twiegand@ieee.org).

images using electroencephalography (EEG) [7] and common spatial patterns (CSP) [8] for dimensionality reduction. For this, [4] does not take into account the relation of the neural signal to MOS values, but classifies distorted from undistorted images and evaluates classification performance with respect to different distortion magnitudes. This provides strong evidence for the general feasibility of SSVEP-based image quality assessment, but does not allow for assessment of perceived quality as no ranking of different quality levels is obtained. CSP is a supervised method that requires labeled training data. It maximizes the variance ratio between two conditions and is commonly used in classification problems. Being supervised, CSP relies on labeled training data that is usually not given in quality assessment studies. Thus, and in contrast to [4], this study evaluates the correlation between brain responses and MOS values and extends [5], [6] by using spatio-spectral decomposition (SSD) [9], an unsupervised method for dimensionality reduction that applies assumptions on the spectral distributions of signal and noise, and aims at maximizing the signal-to-noise ratio (SNR). For this, SSD is re-formulated in the Fourier domain, rendering it particularly suited for the SSVEP framework. In contrast to [5], [6], SSD provides a rational channel selection not relying on neuroanatomical knowledge or training data. By this, neural components are extracted from the EEG-signal, reflecting the sensory processing underlying image quality perception and significant correlations of the extracted neural features to MOS values are achieved. Based on the L_2 -norm of the extracted SSD components we suggest a basic and simple screening method for rejecting those subjects for who SSD fails. Finally, we show the feasibility of the proposed approach by evaluating the prediction of MOS values from neural responses and show a prediction accuracy comparable to behavioral approaches, although our framework still leaves room for further optimization. In order to ease reproducibility, the experimental design and the signal processing pipeline are described in detail.

The paper begins with a short discussion of related work in Section II. Section III gives an introduction to SSVEP and explains underlying neural mechanisms. Section IV and Section V detail the experimental setup, the suggested signal processing and the adapted method for extracting neural components. Results are presented in Section VI and discussed in Section VII.

II. RELATED WORK

EEG-based approaches to assess the perceived quality of multimedia signals have been studied since recently for different stimulus modalities and experimental paradigms. Most approaches in literature exploit event related potentials (ERP, see Section III). For instance, [10], [11], [12], [10] investigate neural correlates of distorted audio and speech signals based on ERP and find significant correlations. The assessment of perceived quality subject to JPEG distortions using ERP is studied in [13]. In [4], steady state visual evoked potentials (SSVEP, see Section III) are used to classify distorted and undistorted images. SSVEP elicited by distorted images are

related to behavioral responses reported as MOS values on a single sensor basis in [5], [6]. The perception of video quality has been studied using EEG in [14], [15], [16], [17], [18]. In [14] correlations between MOS and averaged ERP-amplitudes measured at a single electrode (CPz) of $r = -0.84$ for video quality are reported. For the combination of audio and video, the correlation increases to $r = -0.87$. In [19], changes in different frequency bands of the neural signal as a response to coding artifacts in 3D video are studied. Here, correlations between band power and MOS of $|r| \approx 0.25$ are achieved on a-posteriori selected channels. The relation of EEG-signals to computer graphic artifacts is evaluated in [20]. Multi-variate techniques have been shown capable of detecting the SSVEP in the visual cortex that is induced by the flickering of 3D shutter glasses even when the flickering frequency is about the threshold of perception [21].

A detailed overview on the EEG-based multimedia quality assessment can be found in [22] and [23].

III. STEADY-STATE VISUAL EVOKED POTENTIALS

Neural processing of sensory stimuli changes the electrical activity of the brain. Those changes are stereotypical to specific stimuli and may be measured as potentials between electrodes attached to different locations on the scalp. This technique is called electroencephalography (EEG) [7]. Amplitudes of evoked potentials are usually very small and lie in the range of several microvolts, buried in the EEG background activity with an amplitude range of tens of microvolts. As evoked potentials are time locked with the stimulus onset, they are also called event related potentials (ERP) and can be resolved against the background activity (and other types of non-phase locked noise, introduced e.g. by eye movements) by averaging the recorded signals across several trials [24].

Brain responses elicited in the visual cortex by periodic visual stimulation are periodic themselves and can be very stable in amplitude and phase. They have therefore been referred to as steady-state visual evoked potentials (SSVEP) [25], [26]. Their frequency content is determined by the frequency of the stimulation, as the response spectrum has narrowband peaks at the stimulation frequency and its integer multiples (harmonics) [25]. The occurrence of activity at harmonics of the stimulation frequency has two reasons: Multiple temporal frequencies in the stimulation (e.g. a square-wave temporal modulation profile) and the nonlinearity of the brain [25]. This allows for the description of SSVEP by the amplitude, phase and scalp topography specific to the respective harmonic. As the signal is only contained in the response components harmonically related to the stimulation frequency, the noise in a recording can be easily estimated by the amplitude of the response component in the frequency bins neighbored to harmonics of the stimulation frequency [27]. Responses in real EEG recordings are contaminated by noise. This stems from additive EEG noise [28] or from artifacts introduced e.g. by movements. However, the fact that the response itself is narrowband, while noise sources are broadband explains the reported high signal-to-noise-ratio (SNR) of SSVEP recordings, relative to broadband ERP-responses [25], [27]. For more

detailed information on the SSVEP and its use in research see [25].

IV. EXPERIMENTAL SETUP

A. Stimuli

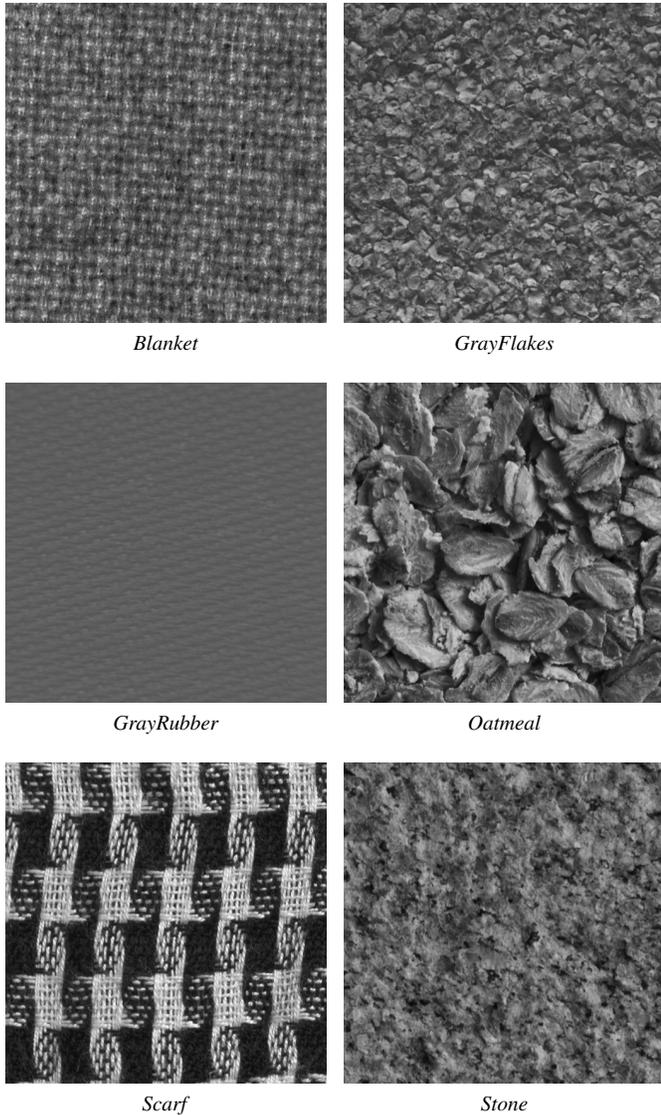


Fig. 1. Texture images used in experiment taken from [29], [30]

In order to make the measurement independent of local image statistics and thus of the current gaze position, six spatially roughly stationary gray-level texture images were chosen from two image databases [30], [29] as the basis for stimulus generation (Fig. 1). The images have a size of 512×512 pixel and are normalized to identical mean luminance. The visual quality of these images was degraded to six different quality levels. Distortions were introduced by coding the original images using the HM10.0 test model [31] of High Efficiency Video Coding standard (HEVC) [32] using *intra only* settings [33]. HEVC offers a flexible quadtree structure for prediction and transform. Statistical redundancies are exploited by block-wise temporal (for video signals) and

spatial linear prediction. The residual signal is transformed block-wise, and coefficients are quantized in the transform domain. Coding artifacts, which are perceived by the human observer as a loss of visual quality, are introduced by the quantization of the transform coefficients. The quantization is controlled by the Quantization Parameter (QP). The distortion levels, mediated by the QP, have been estimated in a pilot study in order to meet roughly similar perceptual qualities with two conditions per texture above the perception threshold (MOS ≈ 8 , see Subsection IV-D), one condition per texture close to the perception threshold and three conditions per texture distributed below the perception threshold. In Fig. 2, examples of the distorted versions of the original image *GrayFlakes* give an impression of the stimuli.

B. Participants

Sixteen participants (seven females and nine males, in the age group 21-46) took part in the experiment. All had normal or corrected-to-normal vision and none of them had a history of neurological diseases. They were all native German speakers or at least with a level of German comprehension of five, on the six level scale of competence laid down by the Common European Framework of reference for Languages [34]. All of them were naïve in respect of video quality assessment studies and were paid for their participation. Each subject was briefed individually about the purpose of the experiment. The study was performed in accordance with the declaration of Helsinki [35] and all participants gave written informed consent.

C. EEG Data Acquisition

EEG was recorded with sampling frequency of 1000 Hz using BrainAmp amplifiers and an ActiCap active electrode system with 64 sensors (both by Brain Products, Munich, Germany). The electrodes used were Fp1,2, AF3,4,7,8, Fz, F1-10, FCz, FC1-6, FT7,8 Cz, C1-6, T7, CPz, CP1-6, TP7,8, Pz, P1-10, POz, PO3,4,7,8, Oz, O1,2. The electrode that in the standard EEG montage is placed at T8 was placed under the right eye and used to measure eye movements. Fig. 3 plots the layout of the electrode positions. All the electrodes were referenced to the left mastoid, using a forehead ground. For offline analyses, electrodes were re-referenced to linked mastoids. All impedances were kept below $10 \text{ k}\Omega$.

D. Stimulus Presentation

As the experiment consisted of two parts, one addressing conventional psychophysical assessment of perceived quality in terms of MOS, the other addressing neurophysiological assessment, two different forms of stimulus presentations were used. However, the presentation environment was identical for both parts of the experiment: The stimuli were shown on a 23" screen (Dell U2311H) with a native resolution of 1920×1080 pixels at a refresh rate of 60 Hz. The screen was normalized according to the specifications in ITU-R Recommendation BT.500 [2]. The size of the images in the behavioral part of the experiment was the same as in the videos. The viewing distance was 110 cm, in compliance with specifications in the

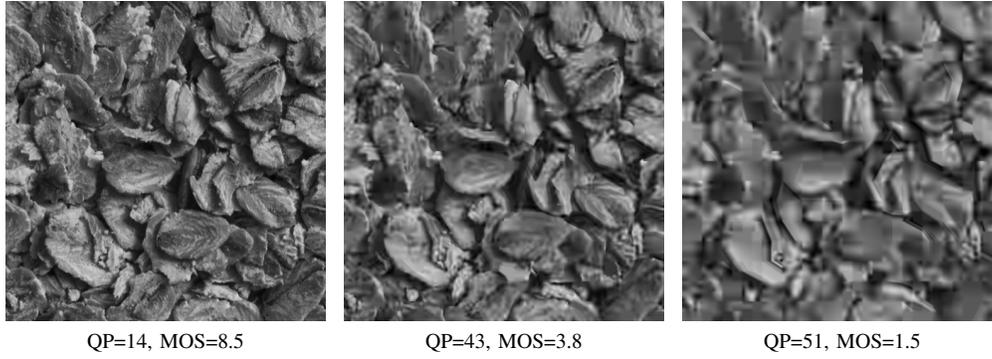


Fig. 2. Distorted images used in experiment exemplified for texture *Oatmeal*.

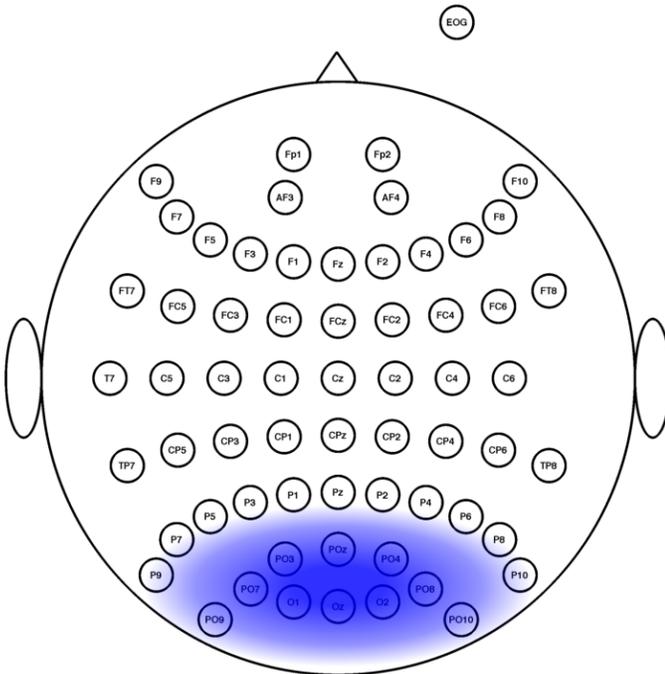


Fig. 3. Electrode layout. The EOG electrode is placed below the right eye. The blue shaded area indicate the approximate location of the visual cortex projected to the scalp.

ITU-T Recommendation P.910 [3], and the stimuli resolution was 512×512 pixels (128×128 mm), which corresponds to 7.15° visual angle. Subjects sat in front of the display in a dimly lit room. Between the two parts of the experiment subjects had a short rest and were provided small snacks and drinks.

1) *Behavioral Part*: In the psychophysical part of the experiment, participants evaluated the perceived quality of the textures following a Degradation Category Rating (DCR) procedure using Simultaneous Presentation (SP) [2]. Image pairs were presented simultaneously side-by-side with the distorted test image on the left hand side and the undistorted reference image on the right hand side within a 50% gray background. Each image pair was presented in the display for 10 s. Each stimulus presentation was followed by a voting during which the observer rated the impairment of the test

image in relation to the reference image using a rating scale implemented by a slider. A nine-grade degradation scale [3] was used where the ratings 1, 3, 5, 7 and 9 corresponded to the semantic annotations *Very annoying*, *Annoying*, *Slightly annoying*, *Perceptible, but not annoying* and *Imperceptible*, respectively. In this scale, grade 8 corresponds to the psychophysical perception threshold of the impairment [3]. Each stimulus pair was presented 3 times in order to obtain statistical significance. Learning effects were reduced by including a training session in which 12 stimuli were presented at the beginning of each session. Stimuli presented during the training session were not included in the statistical analysis of the test results.

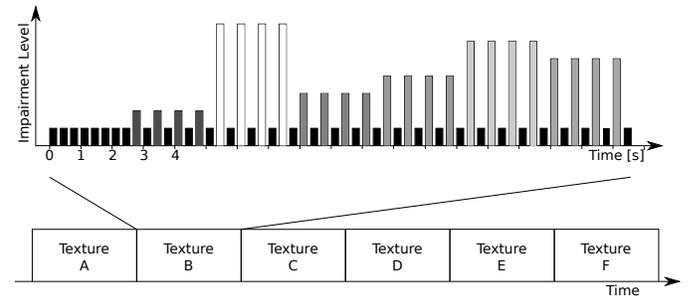


Fig. 4. Structure of the temporal presentation of stimuli, here for one base image. Distorted and undistorted images of one base image are presented alternating at a constant frame rate of $f_r = 3$ Hz. Each distortion was presented for 4 alternation periods, the order of presented distortion levels was randomized in each trial. The first 8 images only comprised undistorted images.

2) *Neurophysiological Part*: In this part of the experiment, in order to elicit SSVEP, distorted images have been presented in periodic alternation with undistorted images at a frame rate of $f_r = 3$ Hz. This corresponds to a stimulation frequency $f_{stim} = 1.5$ Hz in terms of SSVEP paradigms [25].

Stimulus presentation was structured in consecutive texture blocks (see bottom panel of Fig. 4). Within one texture block, alternations have been repeated four times for each distortion magnitude, resulting in a presentation of 8 frames per distortion magnitude. Each texture block started with the presentation of 8 undistorted frames. Thus, a texture-specific block consisted of 56 frames (6 distortion levels \times 4 alternations \times $(1$ distorted frame $+ 1$ undistorted frame) $+ 8$ undistorted frames).

8 undistorted frames). The construction of such a texture-specific block is sketched in the top panel of Fig. 4. The order of the distortions was randomized for each texture and each trial and the order of the textures was randomized for each trial in order to avoid adaptation or memory effects.

The presentation of one trial for all reference images and all distortion levels began with the presentation of a fixation cross and took about 117s in total. Trials were presented in 3 blocks, consisting of 20, 15, and 16, respectively, runs of trials, resulting in 51 trials per condition. The 3 blocks were divided by short breaks of about 10 min.

V. METHODS AND DATA ANALYSIS

A. Analysis of Psychophysical Data

In psychophysical tests, some observers might give inconsistent responses that can distort the result of the test. Those observers can be identified by screening and should be rejected for further analysis as recommended in [2]. Mean opinion scores (MOS) are obtained by averaging condition-wise over the ratings reported by individual observers.

B. Preprocessing of EEG Data

EEG signal was bandpass filtered from 0.5 Hz to 40 Hz using a zero-phase Chebyshev filter of order 8 (3 dB of ripple in the passband and 40 dB of attenuation in the stopbands) and downsampled to 90 Hz. After eye-movement regression, the EEG was re-referenced to the common average of all electrodes.

1) *Eye-Movement Regression*: Let $x_k(t)$ denote the EEG recording on one sensor k and $\mathbf{x}(t) \in \mathbb{R}^K$ represent the recorded signal of K sensors at a time point t . Horizontal eye movements are contained in the difference signal of the sensors F9 and F10 $x_{hor}(t) = x_{F9}(t) - x_{F10}(t)$, vertical eye movements and blinks in the difference between signals measured at electrodes Fp2 and EOG $x_{ver}(t) = x_{Fp2}(t) - x_{EOG}(t)$. By combining $x_{hor}(t)$ and $x_{ver}(t)$ to $\mathbf{x}_{eye} = [x_{hor}(t), x_{ver}(t)]^T$ we define Σ_{eye} as the covariance matrix of $\mathbf{x}_{eye}(t)$, Σ_x as the covariance matrix of $\mathbf{x}(t)$ and $\Sigma_{x,eye}$ as the cross-covariance of $\mathbf{x}_{eye}(t)$ and $\mathbf{x}(t)$. This leads to a *backward model* relating the sensor activity to the underlying originating sources $\mathbf{W} = \Sigma_x^{-1} \Sigma_{x,eye}$ [36]. The forward model, relating the source activity to the observed sensor activities is then given as $\mathbf{A} = \Sigma_x \mathbf{W} \Sigma_{eye}^{-1}$ [36], where Σ_x and Σ_x^{-1} (from \mathbf{W}) cancel out. Interferences of eye motion can now be regressed out from the recorded signal as $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{A} \mathbf{A}^\# \mathbf{x}(t)$ [37], [38], where $\#$ denotes the pseudo-inverse of a matrix. For further processing and analyses, data measured at EOG is neglected.

Note that for the sake of readability, although eye artifacts are regressed out, the recorded data $\mathbf{x}(t)$ is always assumed to be free of eye movement artifacts in the following. Therefore, $\tilde{\mathbf{x}}(t)$ is denoted as $\mathbf{x}(t)$.

2) *Epoching*: The EEG data recorded for each subject is subdivided into epochs ranging from 666.67 ms to 2666.67 ms. Thus, the neural data of each epoch is the signal recorded after the presentation of the second frame of each trial and contains the signal measured for the duration of 3 periods (see

Fig. 4). This reduces the influence of transient components of responses to the stimulus onset [25].

3) *Artifact Rejection*: EEG epochs that contained a large percentage (more than 20%) of data samples exceeding a threshold of 25 μV were excluded as artifacts on a sensor-by-sensor basis. Typically, these epochs were associated with strong eye movements, blinks or other body movement that could not be regressed out.

C. Feature Extraction

Fourier transform is applied epoch-wise to the recorded EEG data. As the sampling frequency ($f_s = 90\text{ Hz}$) is an integer multiple of the stimulation frequency ($f_{stim} = 1.5\text{ Hz}$) and the epoch length of 2 s allows for an integer number of stimulation periods per epoch (in this case 3 periods), no windowing is applied to the data, in order to avoid sidebands [39].

D. Dimensionality Reduction

A common method for the analysis of EEG data and dimensionality reduction is to find a spatial filter \mathbf{W} that projects the sensorwise measurement $\mathbf{x}(t)$ to a new subspace containing the spatial components $\mathbf{y}(t) = \mathbf{W}^T \mathbf{x}(t)$ [36], [40]. As in Section V-B1, \mathbf{W} is found by optimizing $\mathbf{y}(t)$, given $\mathbf{x}(t)$, in regard to a specific criterion. The columns $\mathbf{w}_i \in \mathbb{R}^K$ of $\mathbf{W} \in \mathbb{R}^{K \times K}$ contain the filters of the specific components i . Thus, the time course of the i th spatial component can be calculated as $y_i(t) = \mathbf{w}_i^T \mathbf{x}(t)$. Accordingly, with $\mathbf{A} = (\mathbf{W}^{-1})^T$, the column $\mathbf{a}_i \in \mathbb{R}^K$ of $\mathbf{A} \in \mathbb{R}^{K \times K}$ contain the spatial activity patterns of the respective component i [36].

The signal-to-noise-ratio (SNR) of the components $\mathbf{y}(t)$ was suggested as an optimization criterion for a technique called spatio-spectral decomposition (SSD) [9]. SSD extracts components of neural oscillations by maximizing the power in one frequency band and, simultaneously, minimizing the power in another frequency band. For this the SNR is defined as

$$\text{SNR} = \frac{P_s(f)}{P_n(f)}, \quad (1)$$

where $P_s(f)$ and $P_n(f)$ are power spectra of a recorded EEG signal bandpass filtered according to the assumed signal frequency band (leading to $\mathbf{x}_s(t)$ and its power spectrum $P_s(f)$) and to the assumed noise frequency band (leading to $\mathbf{x}_n(t)$ and its power spectrum $P_n(f)$). With a spatial filter \mathbf{W} and the projection $\mathbf{y}(t) = \mathbf{W}^T \mathbf{x}(t)$, the SNR of $\mathbf{y}(t)$ can be defined by

$$\text{SNR}(\mathbf{W}) = \frac{\mathbf{W}^T \Sigma_s \mathbf{W}}{\mathbf{W}^T \Sigma_n \mathbf{W}} \quad (2)$$

with Σ_s and Σ_n being the covariance matrices of the bandpass filtered signals $\mathbf{x}_s(t)$ and $\mathbf{x}_n(t)$.

Maximizing SNR for \mathbf{W} leads to the generalized eigenvalue problem

$$\Sigma_s \mathbf{W} = \mathbf{D} \Sigma_n \mathbf{W}. \quad (3)$$

The entries of \mathbf{D} contain the generalized eigenvalues and can be interpreted as the amount of SNR projected to a specific

component. The solution of Eq. 3 leads to \mathbf{W} and thus to SNR optimized spatial components. For details on SSD and its application on signals in the time domain, please refer to [9], [40].

With some adaptation, SSD can directly exploit the narrowband property of SSVEPs. The SNR of SSVEPs can be approximated sensor- and harmonic-wise as

$$\text{SNR} = \frac{P_s(f_h)}{P_n(f_h)} \approx \frac{P(f_h)}{0.5(P(f_h - \Delta f) + P(f_h + \Delta f))}, \quad (4)$$

where $P(f_h)$ denotes the power of the harmonic component at one particular harmonic frequency f_h and, with frequency resolution Δf , $P(f_h \pm \Delta f)$ denotes the power of the spectrally neighbored components. Exploiting the unitary property of the Fourier transform, with $X_k(f; e)$ being the Fourier transform of the EEG signal from an epoch e at a sensor k and with K sensors in total, we find the covariance matrices in Eq. 3 as

$$\Sigma_s = \begin{bmatrix} C_{0,0}^s & C_{0,1}^s & \cdots & C_{0,K-1}^s \\ C_{1,0}^s & C_{1,1}^s & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ C_{K-1,0}^s & \cdots & \cdots & C_{K-1,K-1}^s \end{bmatrix} \quad (5)$$

and

$$\Sigma_n = \begin{bmatrix} C_{0,0}^n & C_{0,1}^n & \cdots & C_{0,K-1}^n \\ C_{1,0}^n & C_{1,1}^n & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ C_{K-1,0}^n & \cdots & \cdots & C_{K-1,K-1}^n \end{bmatrix} \quad (6)$$

with

$$C_{i,j}^s = \sum_e X_i(f_h; e) X_j^*(f_h; e) + X_i(-f_h; e) X_j^*(-f_h; e) \quad (7)$$

$$C_{i,j}^n = \sum_e X_i(f_h - \Delta f; e) X_j^*(f_h - \Delta f; e) + X_i(-f_h + \Delta f; e) X_j^*(-f_h + \Delta f; e) + X_i(f_h + \Delta f; e) X_j^*(f_h + \Delta f; e) + X_i(-f_h - \Delta f; e) X_j^*(-f_h - \Delta f; e). \quad (8)$$

Within a SSVEP paradigm, a filter \mathbf{W} can be found for every harmonic of the stimulation frequency.

After solving Eq. 3, we normalize \mathbf{W} columnwise.

VI. RESULTS

A. Behavioral Data

After screening, no participant had to be rejected on basis of the behavioral data. In Fig. 5 the regression of the perceived quality in terms of MOS values is plotted against the QP used for compressing the images. Regression was performed by using a 3-parameter logistic function $\text{MOS}(\text{QP}) = \frac{\beta_1}{1 + e^{-\beta_2 \cdot (\text{QP} - \beta_3)}}$. Vertical bars represent the 95% confidence interval. The perception threshold at MOS=8 [2] is indicated by a horizontal line. The two highest quality

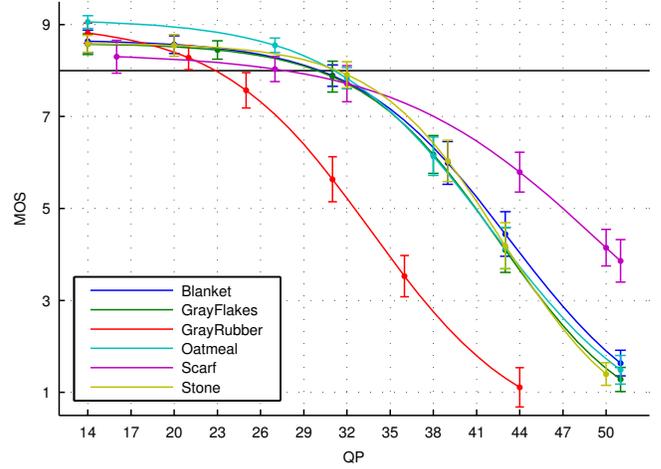


Fig. 5. MOS values obtained in the behavioral part of the experiment for all 6 textures. The distortion perception threshold around MOS = 8 is indicated by the horizontal black line. Vertical errorbars denote confidence intervals of MOS values of the specific condition.

levels are above the perception threshold for each texture. For different textures we tried to introduce distortions at quality levels that can mostly be considered as being perceptually roughly equal. The resulting MOS values are mostly very close and the confidence levels overlap for the most instances of equal quality levels¹. Note the extreme cases, where the horizontally and vertically oriented structure of *Scarf* allows for a relatively good representation by separable DCT and the high contrast is able to mask quantization noise, the rather flat diagonal structure of *GrayRubber* can not be captured by the DCT and structure vanishes due to quantization [41].

B. Neurophysiological Data

As summarized in Section III, a periodic visual stimulation elicits responses consisting of frequency components that are harmonically related with the stimulation frequency f_{stim} . Fig. 6 shows the time courses and the amplitude spectra of the full epoched neural signal measured at Oz electrode and averaged over all trials and textures for subject VPik and for different distortion levels. The distortion level is increasing from top to bottom. The increase of distortion magnitude in the images presented at stimulation frequency f_{stim} triggers an increase of neural processing at f_{stim} and its harmonics. Thus, as the time courses in Fig. 6 show, the EEG signal becomes more and more modulated by the stimulation frequency and its harmonics. This modulation can be quantified directly in the spectral domain. Here, the modulation is represented by increased amplitudes in the spectral components of the harmonics. Although being expected for all harmonics, this behavior is most evident for the even harmonics $f_2 = 3$ Hz and $f_4 = 6$ Hz. Note that it becomes less conclusive for $f_6 = 9$ Hz as the SSVEP is buried in the alpha band of the neural signal. Similar information can be obtained by

¹The maximal QP in HEVC is QP=51. Therefore the quality of *Scarf* could not be reduced any further.

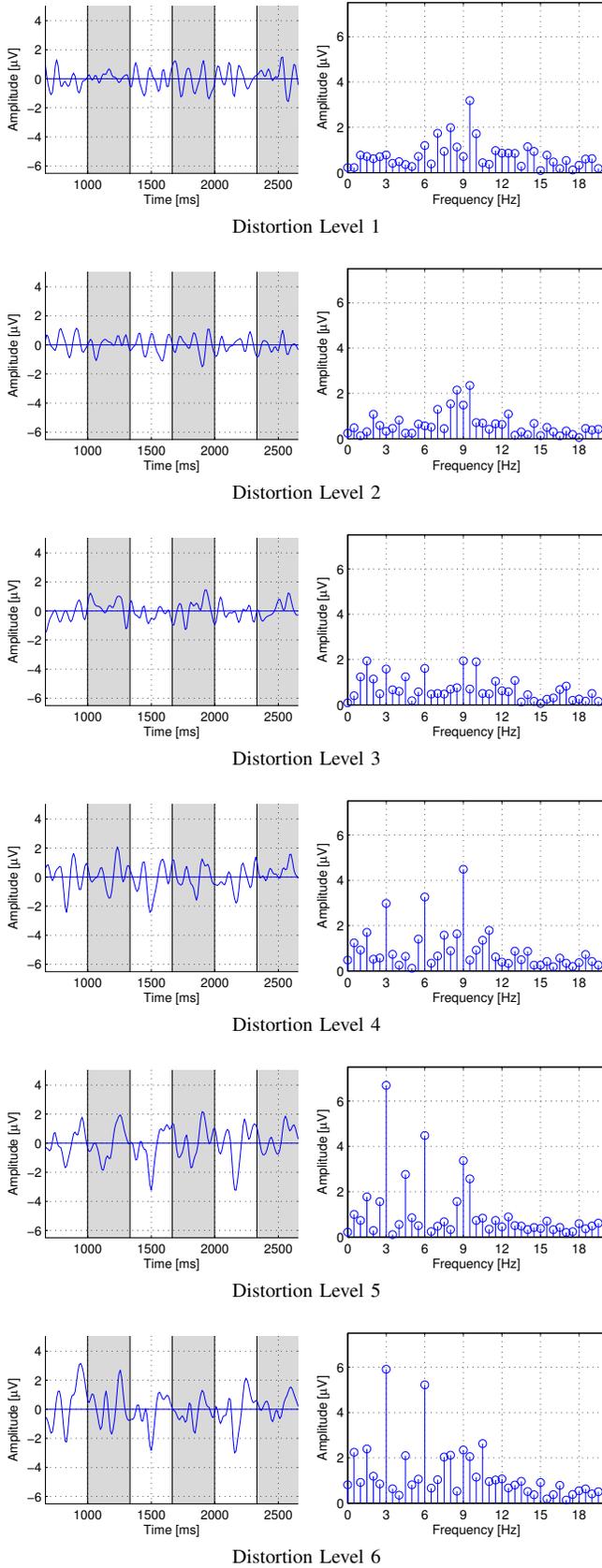


Fig. 6. Epoched neural signals of participant VPik, measured at electrode Oz and averaged over all textures and trials at different distortion levels. Distortions are increasing from top to bottom row. In the left column the time courses are shown, where gray shadings indicate the presentation period of a distorted image. In the right column, each row shows the amplitude spectra of the entire respective averaged epoch. Clearly, the harmonics become more clearly visible in the spectra (right column).

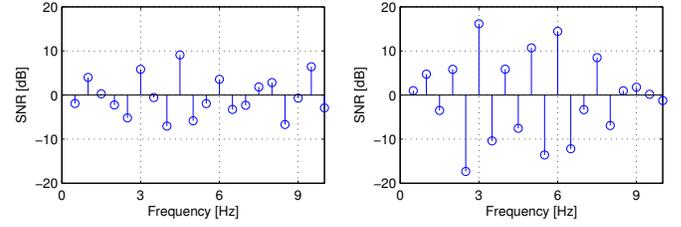


Fig. 7. Signal-to-noise-ratio for the EEG signal measured at electrode Oz (left) and measured for the first component of the SSD optimized for $f_4 = 6$ Hz (right) for subject VPia averaged over all textures for distortion level 4.

analyzing the signal collected at other electrode positions as well (not shown). However, although SSVEPs are expected to be elicited predominantly at electrode positions covering the visual cortex (see blue shaded area in Fig. 3), brain anatomy is strongly variant across humans and cap positions are not perfectly aligned in every experiment. Thus, optimal electrode positions are generally unknown for each measurement.

C. Spatio-Spectral Decomposition

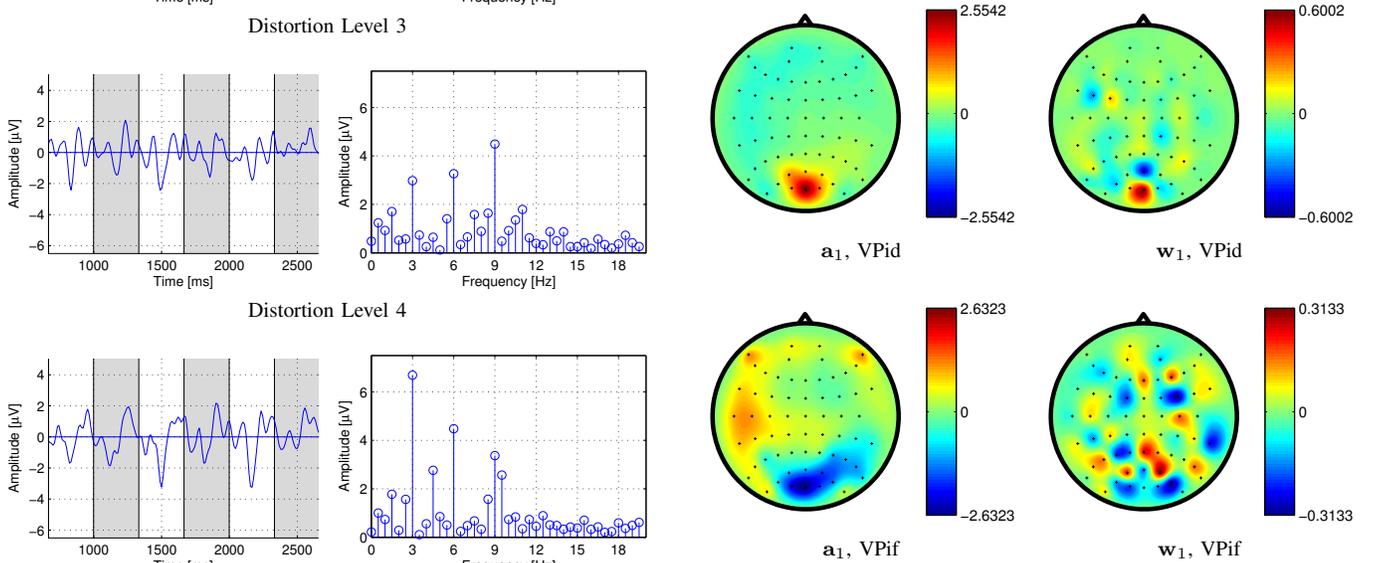


Fig. 8. SSD activation patterns (left column) and SSD filters (right column) for different subjects (from top to bottom: VPid, VPif).

SSD [9] aims at finding a linear combination of sensors that maximize the signal ratio. For this study, optimization is done for each individual subject based on the 4th harmonic $f_4 = 4 \cdot f_{stim} = 6$ Hz of the EEG data, exploiting all epochs and thus not discriminating between different textures or distortion magnitudes. Epoching the EEG data to segments of length of 2000 ms leads to $\Delta f = 0.5$ Hz. The effect of SSD in terms of SNR improvement (see Eq. 4) is exemplified for subject VPia and distortion level 4: Fig. 7 compares the SNR of the signal measured at electrode position Oz to the SNR of the first SSD component. It shows that not only the SNR corresponding to the 4th harmonic $f_4 = 6$ Hz is

improved, but also the SNR corresponding the 2nd harmonic $f_2 = 3$ Hz. Interestingly, the SNR corresponding to the first harmonic $f_1 = 1.5$ Hz and the 3rd harmonic $f_3 = 4.5$ Hz is effectively decreased, although odd harmonics of the SSVEP carry sensory information as well [25].

An impression about the resulting activation patterns \mathbf{A} and filters \mathbf{W} is given in Fig. 8 for the subjects VPid and VPif. Note that the eigenvectors of a matrix are ambiguous regarding their sign, which is why the values of \mathbf{A} are approximately flipped for these two particular subjects. However, the activation pattern show the neurophysiological plausibility of the revealed filters as the highest activation can be found in the area of the visual cortex, most pronounced around electrode Oz, but also stretching to other electrode locations.

D. Relating Neurophysiological to Behavioral Data

Fig. 9 shows the correlations between the average neural signal and the MOS values for all subjects and the grand average (GA). The neural signal is considered in three ways: as the amplitude of the 4th harmonic $f_4 = 6$ Hz of *a*) the signal measured at location Oz (Oz); *b*) the first component of the SSD optimized for the SNR on f_4 (SSD₁); and *c*) the sum of the amplitudes of the first two components of the SSD optimized for the SNR on f_4 (SSD₁ + SSD₂). Significant correlations ($p < 0.05$) are indicated by \circ . For all subjects, the correlation between the signal recorded at Oz and MOS and, except for subject VPir, the correlation between the first SSD component and the sum of the first two SSD components are significant. For most subjects, with the exception of VPig and VPir, the first component of SSD provides a higher correlation to MOS than the signal from Oz. Statistical significance for the differences of the correlations with regard to the correlation of the Oz signal is tested using Steiger's Z-test [42]. In Fig. 9 significant difference in correlations are indicated by $+$ ($p < 0.05$). For 10 out of 16 subjects the increase in correlation by SSD is significant. Although SSD increases correlation for the grand average (from $r_{Oz} = -0.89$ to $r_{SSD_1} = -0.91$), this increase is not statistically significant if all subjects are considered. For the sum of the amplitudes of the two first SSD component, the results becomes less conclusive, as the correlation is further increased for some subjects (eg. VPia or VPic), but decreases for other subjects (e.g. VPal or VPif). For the grand average, the correlations drops even below the correlation achieved in single channel analysis.

The direct relation of the amplitude of the 4th harmonic of the first component of the SSD to the MOS values is shown in Fig. 10.

E. Differences between Subjects

The results for subjects VPig and VPir in Fig. 9 show that SSD, being unsupervised, is not always successful in extracting the neural components that reflect perceived quality. Cases like that are to be expected as they reflect the biological variance among participants. For subject VPig, the correlation can be re-increased by taking the 2nd SSD component into account as well.

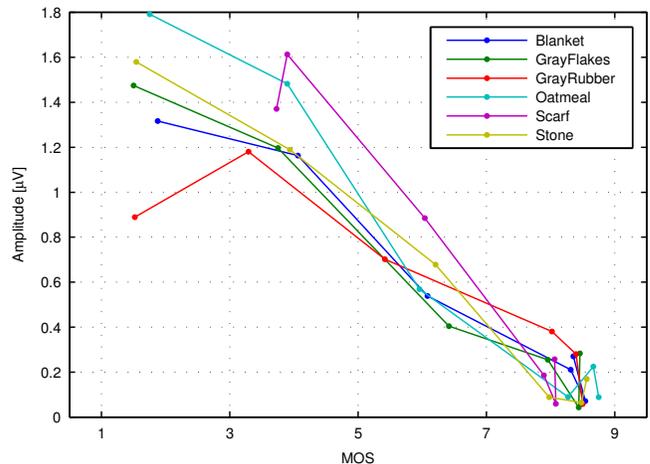


Fig. 10. MOS values vs. grand average of amplitudes of the 1st SSD components at the 4th harmonic frequency component.

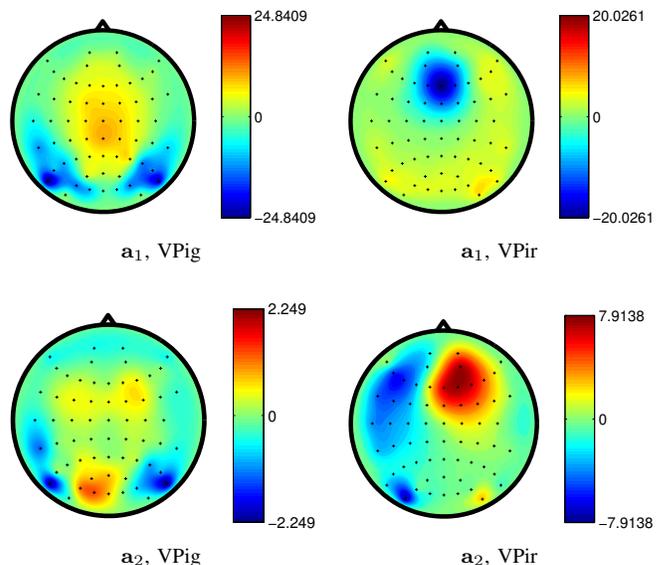


Fig. 11. Activation patterns of the 1st and 2nd SSD components of subjects VPig and VPir. For subject VPir, SSD fails at extracting components activated on the visual cortex.

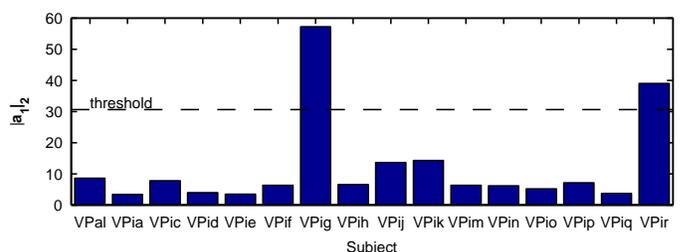


Fig. 12. Subjectwise L_2 -norm of 1st SSD-component. The horizontal dashed line denotes the upper outer fence of all subject's L_2 -norms.

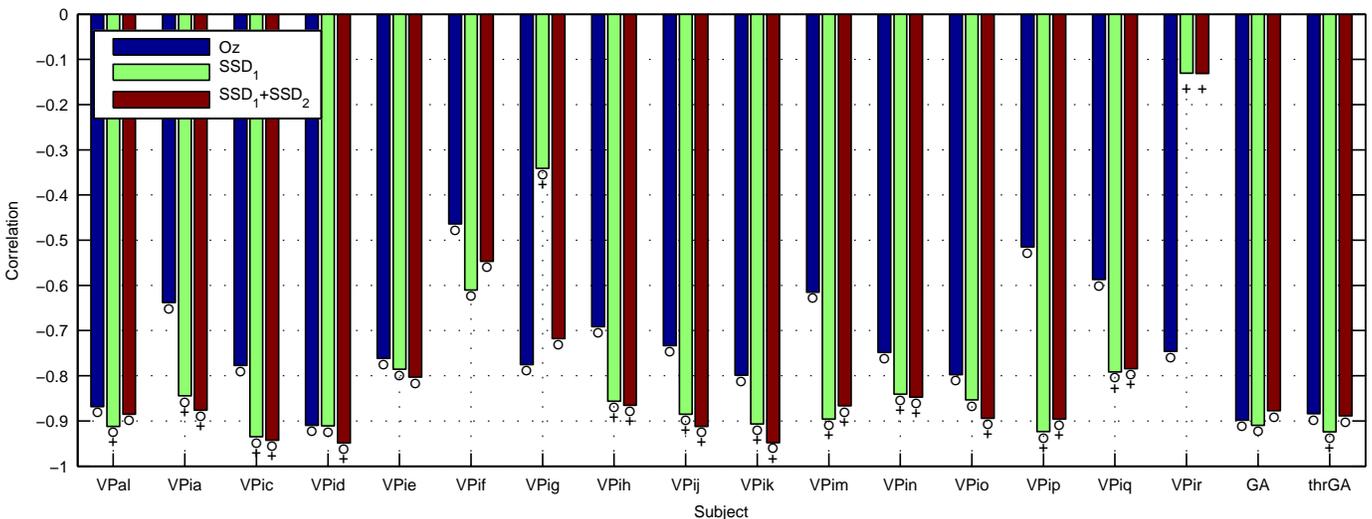


Fig. 9. Correlations between MOS values and texture-wise averaged neural signal for all subjects (VPal-VPir), averaged over all subjects (GA) and averaged over all remaining subjects after screening by thresholding (thrGA). Significant correlations are indicated by \circ , significant increase of correlation of the SSD components with respect to the Oz signal is indicated by $+$ ($p < 0.05$).

The activation patterns for these subjects are shown in Fig. 11. Evidently and in contrast to the results shown in Fig. 8, the activation patterns of the 1st component do not focus on any activation in the visual cortex. This explains the drop in correlation for both subjects, when applying SSD. The 2nd component does represent activity in the visual cortex for subject VPig, but not for subject VPir. For subject VPig, this explains the improved correlation achieved by taking into account the 2nd SSD component as well (see Fig. 9). For VPir, SSD fails at extracting physiologically meaningful components, and thus the correlation stays low even if two SSD components are taken into account.

SSD distributes the activation from channels to components. For failing SSD (Fig. 11) the distribution of activation among the components is different to successful SSD (Fig. 8). Thus, L_2 -norms of the activation pattern of the first SSD component differ significantly between subjects for which SSD fails (VPif and VPir) and subjects for which SSD is able to extract correct components as shown in Fig. 12. This allows to define a measure for screening subjects with regard to successful SSD, employing the the upper outer fence based of the interquartile range [43]: With the threshold $\text{thr} = Q_{75\%} + 3 \cdot (Q_{75\%} - Q_{25\%})$ and $Q_{75\%}$ and $Q_{25\%}$ being the first and third quartile of all subject's $\|\mathbf{a}_1\|_2$ we reject all subjects for which $\|\mathbf{a}_1\|_2$ exceeds thr. Fig. 12 shows $\|\mathbf{a}_1\|_2$ for all subjects. Subjects VPif and VPir can be clearly detected by exceeding thr indicated by the dashed horizontal line. Excluding VPif and VPir increases the overall correlation of the grand average to $r_{SSD_1} = -0.93$. In contrast to the correlation of the grand average considering all subjects, the increase of correlation for the screened grand average is statistically significant with $p < 0.05$ (Fig. 9).

F. Predicting the MOS from the Individual's Neural Signal

The high correlations reported in Subsection VI-D suggest to use a linear model to predict the MOS values from the

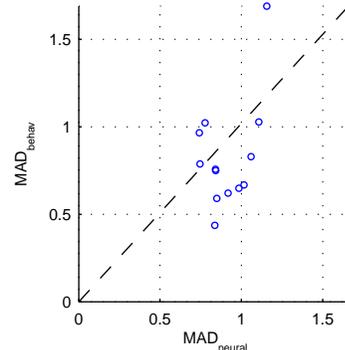


Fig. 13. Residuals of (leave-one-subject-out) MOS from individual subject's opinion (MAD_{behav}) vs. residual from predicting (leave-one-subject-out) MOS from individual subject's neural signal using a linear model cross-validated from all other subject's neural signal (MAD_{neural}).

neural responses. We evaluate the prediction performance subject-wise by leave-one-out cross-validation. Based on a linear model

$$y = \beta_0 + \beta_1 x + \epsilon, \quad (9)$$

with x being the amplitude of the first SSD component at 6 Hz of an individual subject S_i , y being MOS values, β_0, β_1 being the regression coefficients and ϵ the prediction error. The regression coefficients β_0, β_1 are estimated for each subject S_i based on the MOS values pooled for all subjects except S_i and the first SSD components of all subjects except S_i . In order to account for subject-wise differences in amplitude ranges [44], caused e.g. by anatomical differences among the subjects, we normalize the amplitudes of the neural signals subjectwise over all source images and distortion levels to a range between 0 and 1. The prediction performance is quantified subjectwise as MAD_{neural} , the mean absolute difference (MAD) between the MOS predicted from the neural signal of S_i and the observed MOS from all subjects but S_i over all source images and distortion levels. For comparison and

as behavioral counterpart, we evaluate the prediction of the cross-validated MOS (pooled from all subjects but S_i) from the behavioral rating of S_i . The prediction performance is quantified as MAD_{behav} , the MAD between subject S_i ratings and the MOS pooled from all subjects except S_i over all source images and distortion levels. Fig. 13 shows the prediction performances of the proposed approach and the behavioral approach in terms of MAD_{neural} and MAD_{behav} . Each circle represents the prediction performances of one subject. The dashed line indicates identity of the two prediction performances. Although most points are located slightly below the dashed line, and by that suggesting slightly higher accuracy of the behavioral approach, assuming Gaussian distribution of MAD_{neural} and MAD_{behav} t-test [42] reveals that the means of MAD_{neural} and MAD_{behav} are statistically equal ($p < 0.05$). Giving up the assumption of Gaussianity, signed-rank test [42] also shows statistical equality of the medians of MAD_{neural} and MAD_{behav} ($p < 0.05$).

VII. DISCUSSION

We presented a neurophysiological approach to image quality assessment exploiting SSVEP. For analysis, we adapted the SSD technique specifically for SSVEP data, used it to extract physiologically meaningful neural components from the recorded data and showed that for most of the subjects these components have a statistically significantly higher correlation to behavioral responses than the signal recorded at Oz. By using SSD, we overcome the problem of channel selection in SSVEP-based image quality assessment. This paper did not present a final solution for objective and reliable assessment of video quality assessment, but we showed that with the presented method high correlations of the extracted neural signal with MOS values are achieved and that the proposed method is feasible to predict MOS values with an accuracy comparable to behavioral approaches.

The SSVEP approach is able to achieve a significantly higher SNR than ERP-based approaches [25], and also the number of trials collected per time is much higher compared to ERP-based approaches. However, so far studies evaluating SSVEP and ERP for image quality assessment have used different sets of stimuli, so a final conclusion is difficult. To understand the differences of the two paradigms on the next level, it will be important to establish a similar set of stimuli and then conduct future experiments to allow a precise comparison and identify strengths and weaknesses of the two approaches.

In order to arrive at a real-world solution to quality assessment, this paper is limited in following respects and raises several challenges for future work:

Most important, subjects sensitive to photonic flicker might suffer not only from headache, but even seizures could in principle be introduced by the presentation of a flickering stimulus if the subject is suffering from epilepsy [45]. In an SSVEP-based quality assessment study this must be prevented by identifying and excluding affected subjects from the experiments.

Eliciting SSVEPs relies on temporally highly precise and alternating stimulus presentation. This is a clear limitation

of the proposed approach and renders SSVEP-based quality assessment 'in the wild' a very hard challenge.

As for all ERPs-based approaches as well, the presented SSVEP focuses on the perception of quality change (introduced by the alternation between undistorted and distorted images) rather than quality perception per se. This bounds the approach to the full reference domain and might limit its applicability to only certain real-world applications.

Although our results indicate clear feasibility of the proposed approach, future work should consider to study the influence of low level image statistics such as luminance or contrast systematically on the prediction performance.

In this study, subjects for which only a low correlation between neural signal and MOS values was obtained (e.g. VPif, see Fig. 9) could not be identified by conventional psychophysical screening [2]. On the recorded data, a simple screening method based on interquartile ranges [43] was shown to be useful for identifying subjects for which SSD failed. By that the performance of the proposed method in terms of correlation is statistically significantly increased. Replicability of this screening approach will have to be evaluated on other recordings, other subjects and for paradigms other than SSVEP. Thus, in order to allow for a real-world application of neurophysiological quality assessment methods, analogously to psychophysical methods, appropriate screening techniques will need to be further evaluated. Identifying and excluding people for which BCI methods fail [46], [47], [48] is able to boost performance of EEG-based methods. As an example, [4] shows a negative relation between EEG-based distortion detection and the power in the α -band (7.5 Hz to 12.5 Hz) and argues that the α -activity interferes with the processing of the visual information, while a state of high cortical excitability is reflected by decreased α -activity. For future work, this observation can serve as a starting point to study screening methods for EEG-based quality assessment. Also, identifying 'high performing' subjects may reduce the number of subjects necessary for EEG-based quality assessment studies.

Results presented in this study are based on averages across trials. Besides identifying the number of subjects, for real world applications it is crucial to identify the number of trials necessary for reliable quality assessment studies and in the optimal case allow for single trial quality assessment. The question regarding the number of subjects and the number of trials can be treated analogously to psychophysical approaches [2]. However, in order to move towards applying the proposed method generically in image quality assessment studies, parameters of the experimental design need to be optimized. Important factors potentially driving the performance of the approach include the stimulation frequency used for eliciting the SSVEP and the dimensionality reduction method. In the presented study, the stimulation frequency was set to $f_{\text{stim}} = 1.5$ Hz. It is known that for specific cognitive tasks there are optimal stimulation frequencies [49], [50]. Future work should evaluate if such an optimal stimulation frequency also exists for image quality assessment. By using an optimal stimulation frequency, the duration of the experiments might be shortened. Further, the results in this paper are based only on the 4th harmonic component, while SSVEPs are

elicited on several harmonic frequencies. It has to be evaluated how other frequency components can be integrated into an analysis framework and exploited simultaneously. This is especially important as it was shown that different harmonic components represent different neural processing [51]. As different aspects of neural processing usually implies different neural sources and thus different activation topographies [52], this may explain the drop of SNR that is observed for odd harmonics of an SSD component that is optimized for even harmonics as shown in Fig. 7. For some subjects, SSD failed to increase the correlation to behavioral responses and to extract neurophysiological plausible components. One reason for that (in contrast to CSP in [4]) is that SSD is an unsupervised channel decomposition technique. Although a strategy for identifying subjects for which SSD fails was proposed, it would be beneficial to enhance the robustness of dimensionality reduction methods e.g. by divergence methods that allow a higher resistance to outlier trials or other noise contamination [53]. For in-lab quality assessment however, e.g. for creating a database of quality annotated images or for exploring some kind of distortion related parameter space, however, SSVEP shows potential to be a feasible paradigm as we showed that the accuracy of the subject-wise prediction of MOS values from the extracted neural responses is statistically equal to the prediction based on behavioral methods even though the proposed methods has room for optimization. Here, the linear prediction model was estimated on the data of all other subjects. However, for application scenarios it would be beneficial to identify a subject-wise model that does not rely on other subjects' responses to predict MOS values directly from individual neural responses.

Future studies may aim at distortion levels close to the perception threshold as this is as desirable operational point for image communication systems. Here, a neurally informed quality assessment procedure might help to complement conventional behavioral methods, taking into account the heteroskedastic noise characteristics at the edge of perception [54].

In this study, we used SSVEP to assess the perception of quality of images containing texture only in order to ease the experimental setup. Conceptually there is no reason to limit the proposed approach to this class of stimuli. Thus, it will be interesting to study experimentally whether the proposed approach is also a feasible method to assess perceived quality of complex natural images. SSVEPs have also been used to assess motion perception [25]. Following this line, the feasibility of SSVEP to assess video quality could be evaluated in an extension of the presented experimental setup.

The extra preparation time ($\approx \frac{1}{2}$ h) required for the setup of the EEG system might eliminate the benefits of EEG measurements, but a new generation of dry electrode-based EEG caps has the potential to shorten the preparation time drastically. It is recommended for psychophysical experiments not to last longer than 30 minutes, in order to prevent the subject from becoming unreliable in their behavioral responses due to fatigue or boredom. In EEG-based experiments in contrast, no response has to be given by the subjects and it is not known yet what the limits in terms of duration are; in

cognitive neuroscience length of EEG-based experiments can range between 2-3 hours.

We evaluated and quantified quality related neural correlates based on an SSVEP paradigm. Clearly, several aspects of the presented method need further evaluations and improvements, but we showed that neural signals significantly correlate to perceived quality are elicited and that spatial filtering using SSD increases the correlation for most of the subjects. By this, potentially a less biased and more objective measure of quality perception than obtained with conventional behavioral methods can be established.

REFERENCES

- [1] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
- [2] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Geneva, Switzerland, Tech. Rep., 2012.
- [3] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, Switzerland, Tech. Rep., apr 2008.
- [4] L. Acqualagna, S. Bosse, A. K. Porbadnigk, G. Curio, K.-R. Müller, T. Wiegand, and B. Blankertz, "EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs)." *J. Neural Eng.*, vol. 12, no. 2, p. 026012, 2015.
- [5] S. Bosse, L. Acqualagna, A. K. Porbadnigk, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand, "Neurally informed assessment of perceived natural texture image quality," in *Image Process. (ICIP), 2014 IEEE Int. Conf.* IEEE, oct 2014, pp. 1987—1991.
- [6] S. Bosse, L. Acqualagna, A. K. Porbadnigk, G. Curio, K.-R. Müller, B. Blankertz, and T. Wiegand, "Neurophysiological assessment of perceived image quality using steady-state visual evoked potentials," in *SPIE Opt. Eng. Appl.*, vol. 9599, International Society for Optics and Photonics. International Society for Optics and Photonics, 2015, pp. 959914–959914.
- [7] H. Berger, "Über das Elektroencephalogramm des Menschen," *European archives of psychiatry and clinical neuroscience*, vol. 98, no. 1, pp. 231–254, 1933.
- [8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [9] V. V. Nikulin, G. Nolte, and G. Curio, "A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition," *Neuroimage*, vol. 55, no. 4, pp. 1528–1535, 2011.
- [10] A. K. Porbadnigk, M. S. Treder, B. Blankertz, J. N. Antons, R. Schleicher, S. Möller, G. Curio, and K.-R. Müller, "Single-trial analysis of the neural correlates of speech quality perception." *J. Neural Eng.*, vol. 10, no. 5, p. 056003, oct 2013.
- [11] A. K. Porbadnigk, J.-N. Antons, B. Blankertz, M. S. Treder, R. Schleicher, S. Möller, and G. Curio, "Using ERPs for assessing the (sub) conscious perception of noise," in *Eng. Med. Biol. Soc. (EMBC), 2010 Annu. Int. Conf. IEEE*, vol. 2010. IEEE, jan 2010, pp. 2690–2693.
- [12] J.-N. Antons, R. Schleicher, S. Arndt, S. Moller, A. K. Porbadnigk, and G. Curio, "Analyzing Speech Quality Perception Using Electroencephalography," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 6, pp. 721–731, oct 2012.
- [13] L. Lindemann and M. Magnor, "Assessing the quality of compressed images using EEG," in *2011 18th IEEE Int. Conf. Image Process.*, 2011, pp. 3109–3112.
- [14] S. Arndt, J.-N. Antons, R. Schleicher, S. Moller, and G. Curio, "Using electroencephalography to measure perceived video quality," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 3, pp. 366–376, 2014.
- [15] M. Mustafa, S. Guthe, and M. Magnor, "Single-trial EEG classification of artifacts in videos," *ACM Trans. Appl. Percept.*, vol. 9, no. 3, pp. 1–15, 2012.
- [16] L. Lindemann, S. Wenger, and M. Magnor, "Evaluation of video artifact perception using event-related potentials," *Proc. ACM SIGGRAPH Symp. Appl. Percept. Graph. Vis. - APGV '11*, vol. 0, no. 0, p. 53, 2011.
- [17] S. Scholler, S. Bosse, M. S. Treder, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand, "Toward a direct measure of video quality perception using EEG." *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2619–29, 2012.

- [18] J. Perez and E. Delecquelle, "On the measurement of image quality perception using frontal EEG analysis," in *2013 Int. Conf. Smart Commun. Netw. Technol.*, 2013, pp. 1–5.
- [19] E. Kroupi, P. Hanhart, J.-S. Lee, M. Rerabek, and T. Ebrahimi, "EEG correlates during video quality perception," in *Signal Process. Conf. (EUSIPCO), 2014 Proc. 22nd Eur.* IEEE, 2014, pp. 1–4.
- [20] M. Mustafa and M. Magnor, "ElectroEncephaloGraphics: Making Waves in Computer Graphics Research," *Comput. Graph. Appl. IEEE*, vol. 34, no. 6, pp. 46–56, 2014.
- [21] M. A. Wenzel, R. Schultze-Kraft, F. C. Meinecke, F. Cardinaux, T. Kemp, K.-R. Müller, G. Curio, and B. Blankertz, "EEG-based usability assessment of 3D shutter glasses," *J. Neural Eng.*, vol. 13, no. 1, p. 016003, 2016.
- [22] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek, "Brain-Computer Interfacing for Multimedia Quality Assessment," in *Proc. Work. Brain-Machine Interface Syst. IEEE Int. Conf. Syst. Man, Cybern.* IEEE, 2016.
- [23] U. Engelke, D. Darcy, G. Mulliken, S. Bosse, M. Martini, S. Arndt, J.-N. Antons, K. Chan, N. Ramzan, and K. Brunnstrom, "Psychophysiology-Based QoE Assessment: A Survey," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2017.
- [24] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single-trial analysis and classification of ERP components - A tutorial," *Neuroimage*, vol. 56, no. 2, pp. 814–825, 2011.
- [25] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottreau, and B. Rossion, "The steady-state visual evoked potential in vision research: A review," *J. Vis.*, vol. 15, no. 6, pp. 1–46, 2015.
- [26] D. Regan, "Steady-state evoked potentials," *J. Opt. Soc. Am.*, vol. 67, no. 11, pp. 1475–1489, nov 1977.
- [27] T. Meigen and M. Bach, "On the statistical significance of electrophysiological steady-state responses," *Doc. Ophthalmol.*, pp. 207–232, 1999.
- [28] J. Victor and J. Mast, "A new statistic for steady-state evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, pp. 378–388, 1991.
- [29] G. Kylberg, "The Kylberg Texture Dataset v. 1.0," Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden, External report (Blue series) 35, sep 2011.
- [30] T. Ojala, J. Viertola, S. Huovinen, M. Vision, and M. P. Unit, "Outex - New Framework for Empirical Evaluation of Texture Analysis Algorithms," *Pattern Recognit.*, vol. 1, pp. 701–706, 2002.
- [31] JCT-VC, "Subversion Repository for the HEVC Test Model reference software," 2014.
- [32] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [33] F. Bossen, "Common test conditions and software reference configurations Output," document JCTVC-H1100 of JCT-VC, San José CA, USA, pp. 1–10, 2013.
- [34] D. Little, "The issue: The common European framework of reference for languages: Perspectives on the making of supranational language education policy," *Mod. Lang. J.*, vol. 91, no. 4, pp. 645–655, 2007.
- [35] WMA General Assembly, "Declaration of Helsinki. Ethical principles for medical research involving human subjects." pp. 1–8, 2013.
- [36] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J. D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96–110, 2014.
- [37] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *Neuroimage*, vol. 28, no. 2, pp. 326–341, 2005.
- [38] K.-R. Müller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 165–169, 2003.
- [39] M. Bach and T. Meigen, "Do's and don't's in Fourier analysis of steady-state potentials," *Doc. Ophthalmol.*, vol. 99, no. 1, pp. 69–82, jan 1999.
- [40] S. Haufe, S. Dähne, and V. V. Nikulin, "Dimensionality reduction for the analysis of brain oscillations," *Neuroimage*, vol. 101, pp. 583–597, 2014.
- [41] T. Wiegand and H. Schwarz, "Source Coding: Part II of Fundamentals of Source and Video Coding," *Foundations and Trends® in Signal Processing*, vol. 10, no. 1, pp. 1–346, 2016.
- [42] D. C. Howell, *Statistical Methods for Psychology*. Cengage Learning, 2013.
- [43] J. W. Tukey, *Exploratory data analysis*. Addison-Wesley, 1977.
- [44] H. Strasburger, W. Scheidler, and I. Rentschler, "Amplitude and phase characteristics of the steady-state visual evoked potential," *Applied Optics*, vol. 27, no. 6, pp. 1069–1088, 1988.
- [45] R. S. Fisher, G. Harding, G. Erba, G. L. Barkley, and A. Wilkins, "Photic- and pattern-induced seizures: A review for the epilepsy foundation of america working group," *Epilepsia*, vol. 46, no. 9, pp. 1426–1441, 2005.
- [46] H. I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S. W. Lee, "Predicting BCI subject performance using probabilistic spatio-temporal filters," *PLoS One*, vol. 9, no. 2, 2014.
- [47] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *Neuroimage*, vol. 51, no. 4, pp. 1303–1309, 2010.
- [48] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, and A. Köbler, "Psychological predictors of SMR-BCI performance," *Biological Psychology*, vol. 89, no. 1, pp. 80–86, 2012.
- [49] E. Alonso-Prieto, G. Van Belle, J. Liu-Shuang, A. M. Norcia, and B. Rossion, "The 6 Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex." *Neuropsychologia*, vol. 51, no. 13, pp. 2863–75, nov 2013.
- [50] D.-O. Won, H.-J. Hwang, S. Dähne, K.-R. Müller, and S.-W. Lee, "Effect of higher frequency on the classification of steady-state visual evoked potentials." *Journal of neural engineering*, vol. 13, no. 1, p. 016014, 2015.
- [51] J. Liu-Shuang, J. J. M. Ales, B. Rossion, and A. M. Norcia, "The effect of contrast polarity reversal on face detection: Evidence of perceptual asymmetry from sweep VEP," *Vision Res.*, vol. 108, pp. 8–19, 2015.
- [52] A. Norcia, J. Ales, E. Cooper, and T. Wiegand, "Measuring perceptual differences between compressed and uncompressed video sequences using the swept-parameter Visual Evoked Potential," *J. Vis.*, vol. 14, no. 10, pp. 649–649, aug 2014.
- [53] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.
- [54] A. K. Porbadnigk, N. Görnitz, C. Sannelli, A. Binder, M. Braun, M. Kloft, and K.-R. Müller, "Extracting latent brain states - Towards true labels in cognitive neuroscience experiments," *Neuroimage*, vol. 120, pp. 225–253, 2015.



Sebastian Bosse Sebastian Bosse is with the Machine Learning group and the Image Video Coding group in the Video Coding & Analytics Department of Fraunhofer Heinrich Hertz Institute, Berlin, Germany. He studied Electrical Engineering at RWTH Aachen University, Germany and Polytechnic University of Catalonia, Barcelona, Spain, and received the Dipl.-Ing. degree in Electrical Engineering and Information Technology from RWTH Aachen University, Germany. He was a visiting researcher at Siemens Corporate Research, Princeton, USA and at Stanford University, USA. His major research interests include image and video compression, human visual perception for image communications, and neural correlates of perceived visual quality, as well as signal processing and machine learning.



Laura Acqualagna Laura Acqualagna is a researcher and PhD candidate at the Neurotechnology group, Technische Universität Berlin (TU Berlin), since 2014. She first joined TU Berlin in 2011 as researcher at the Machine Learning group. She received the MSc. in Biomedical Engineering, with focus on Neuroengineering, in 2010 at Università degli studi di Genova (UNIGE). Her research interests are Brain-Computer Interfaces (BCIs), especially in investigating novel BCI paradigms and machine learning techniques for clinical applications

and human-computer interaction.



Wojciech Samek Wojciech Samek received a Diploma degree in computer science from Humboldt University of Berlin, Berlin, Germany, in 2010 and the Ph.D. degree in machine learning from the Technische Universität Berlin, Berlin, in 2014. In the same year, he founded the Machine Learning Group, Fraunhofer Heinrich Hertz Institute, Berlin, which he currently directs. He is associated with the Berlin Big Data Center (BBDC) and was a Scholar of the German National Academic Foundation and a Ph.D. Fellow at the Bernstein Center for Computational

Neuroscience (BCCN) Berlin. He was visiting the Heriot-Watt University, Edinburgh, U.K., and the University of Edinburgh, Edinburgh, from 2007 to 2008. In 2009, he was with the Intelligent Robotics Group, NASA Ames Research Center, Mountain View, CA, USA. His research interests include machine learning, neural networks, robust signal processing, and computer vision.



Klaus-Robert Müller (M'12) has been a professor of computer science at Technische Universität Berlin since 2006; at the same time he has been the director of the Bernstein Focus on Neurotechnology Berlin. He studied physics in Karlsruhe from 1984 to 1989 and obtained his Ph.D. degree in computer science at Technische Universität Karlsruhe in 1992. After completing a postdoctoral position at GMD FIRST in Berlin, he was a research fellow at the University of Tokyo from 1994 to 1995. In 1995, he founded the Intelligent Data Analysis group at GMD-FIRST

(later Fraunhofer FIRST) and directed it until 2008. From 1999 to 2006, he was a professor at the University of Potsdam. He was awarded the 1999 Olympus Prize by the German Pattern Recognition Society, DAGM, and, in 2006, he received the SEL Alcatel Communication Award and in 2014 he was granted the Science Prize of Berlin awarded by the Governing Mayor of Berlin. In 2012, he was elected to be a member of the German National Academy of Sciences-Leopoldina. His research interests are intelligent data analysis, machine learning, signal processing, and brain-computer interfaces.



Anne K. Porbadnigk Anne K. Porbadnigk received her M.S. degree in Computer Science from the University of Michigan, Ann Arbor (USA), in 2006, and the German Diplom in Computer Science from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2009. She spent the academic year 2008/2009 as a Visiting Researcher at Carnegie Mellon University, Pittsburgh, PA (USA). She joined the Machine Learning Laboratory, Berlin Institute of Technology, Berlin, Germany, in 2009, working as a Research Assistant in the Berlin Brain Computer

Interface (BCCI) group. She is also affiliated with the Bernstein Center for Computational Neuroscience Berlin.



Gabriel Curio Gabriel Curio received the Dr. med. degree from Freie Universität (FU), Berlin, Germany, with a thesis about attentional influences on smooth pursuit eye movements. Since 1991, he has been leading the Neurophysics Group at the FU Berlin. Currently, he is a Professor of Neurology and Deputy Director at the Department of Neurology and Clinical Neurophysiology, Campus Benjamin Franklin. He is Founding Co-Director of the Bernstein Center for Computational Neuroscience Berlin and the Berlin NeuroImaging Center, Founding

Member of the Bernstein Focus Neurotechnology Berlin, and Faculty member of the Berlin Excellence School Mind and Brain. His current research interests include integrating the neurophysics of noninvasive electromagnetic brain monitoring with both basic and clinical neuroscience concepts.



Thomas Wiegand Thomas Wiegand is a professor in the department of Electrical Engineering and Computer Science at the Technical University of Berlin and is jointly heading the Fraunhofer Heinrich Hertz Institute, Berlin, Germany. He received the Dipl.-Ing. degree in Electrical Engineering from the Technical University of Hamburg-Harburg, Germany, in 1995 and the Dr.-Ing. degree from the University of Erlangen-Nuremberg, Germany, in 2000. As a student, he was a Visiting Researcher at Kobe University, Japan, the University of California at

Santa Barbara and Stanford University, USA, where he also returned as a visiting professor. He was a consultant to Skyfire, Inc., Mountain View, CA, and is currently a consultant to Vidyo, Inc., Hackensack, NJ, USA. Since 1995, he has been an active participant in standardization for multimedia with many successful submissions to ITU-T and ISO/IEC. In 2000, he was appointed as the Associated Rapporteur of ITU-T VCEG and from 2005-2009, he was Co-Chair of ISO/IEC MPEG Video. The projects that he co-chaired for the development of the H.264/MPEG-AVC standard have been recognized by an ATAS Primetime Emmy Engineering Award and a pair of NATAS Technology & Engineering Emmy Awards. For his research in video coding and transmission, he received numerous awards including the Vodafone Innovations Award, the EURASIP Group Technical Achievement Award, the Eduard Rhein Technology Award, the Karl Heinz Beckurts Award, the IEEE Masaru Ibuka Technical Field Award, and the IMTC Leadership Award. He received multiple best paper awards for his publications. Thomson Reuters named him in their list of 'The World's Most Influential Scientific Minds 2014' as one of the most cited researchers in his field. He is a recipient of the ITU150 Award.

Benjamin Blankertz Benjamin Blankertz received his PhD in mathematics in 1997. After pursuing several studies in music cognition, he started Brain-Computer Interface (BCI) research in 2000. He became chair for Neurotechnology at Technische Universität Berlin in 2012. The Berlin BCI group is known for innovative machine learning approaches in the field of BCI and the development of novel experimental paradigms. This includes, the transfer of BCI technology from the lab to real world applications.