# Brain-Computer Interfacing for Multimedia Quality Assessment

Sebastian Bosse*, Klaus-Robert Müller†‡, *Member, IEEE*,
Thomas Wiegand*§, *Fellow, IEEE*, and Wojciech Samek*, *Member, IEEE*
*Department of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany
† Machine Learning Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany
‡Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea
§Image Communication Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany

***Abstract*—The assessment of perceived multimedia quality is a central research field in information and media technology. Conventionally, psychophysical techniques are used for determining the quality of multimedia signals. Recently, Brain-Computer Interfacing (BCI)-based methods have been proposed for the assessment of perceived multimedia signal quality. In this paper we give an overview over the shortcomings of conventional approaches, present the state-of-the art of BCI-based methods and discuss open questions and challenges relevant to the BCI community.**

## I. Introduction

Balancing resources such as channel bandwidth, system complexity or production costs in order to provide the user with optimal signal quality is a crucial task when it comes to the design of multimedia systems. For most multimedia systems, the ultimate receiver is human and, thus, signal quality should be evaluated in a perceptual relevant manner in order to achieve best user experience. However, the formation of user experience is still not sufficiently understood. For the evaluation of multimedia systems or for testing hypotheses, researchers and engineers in the field of quality assessment (QA) rely on behavioral experiments during which subjects give overt responses to specific stimuli.

Recently, BCI entered a broader scope of definition and monitoring and decoding the mental state of humans became an active research field [1]. As mental states are reflectances of sensation, perception and decision making, this makes BCI a perfect candidate to provide insights into the neural processing of quality experience. It might further ease, refine or even replace conventional behavioral test procedures, and allow for real-time quality monitoring in immersive environments such as virtual reality (VR) without breaking the impression of immersiveness by asking users for their current experience.

This paper aims at providing an overview on BCI-based multimedia quality assessment and identifying challenges in this emerging field that are relevant for the BCI-community. A summary on commonly used behavioral methods and their drawbacks is given in Sec. II. Sec. III outlines approaches to assess the quality of different signal modalities. The paper concludes in Sec. IV with a discussion of open questions and current challenges in BCI-based QA.

## II. Conventional Psychophysical Multimedia Quality Assessment

Traditionally, perceived quality of multimedia signals or systems is assessed in psychophysical subjective tests. In these experiments a human observer is asked to give a judgement on the perceived quality of a multimedia signal presented. Methodologies of such subjective testing are formalized by the International Telecommunication Union (ITU), e.g. for television applications in [2], for multimedia applications in [3], and in [4] for speech quality. These formalizations (or in ITU terminology *Recommendations*) comprise the viewing (such as viewing distance and background illuminations) and/or listening (such as reverberation time or environmental noise) conditions as well as experimental parameters such as stimulus presentation methods, category scales and scale annotations, and processing and reporting of the collected data.

Stimuli can be presented in a single stimulus manner one stimulus at a time, where the subject is asked after each presentation to give his or her rating. This procedure is called absolute category scaling (ACR) [3] and can be performed with or without presentation of a hidden reference (ACR-HR). Stimuli can also be presented in pairs, where the first stimulus is always the undistorted source reference and the second stimulus the distorted source under test. This method is called degradation category scaling (DCR) [3] or double stimulus impairment scaling (DSIS) [2] and provides a higher sensitivity for test stimuli with degradations close to the perception threshold. For the quality assessment of visual media such as images or video following a DCR procedure, the stimuli can also be presented simultaneously. Rating scales for ACR and DCR may be continuous or discrete and different numbers of grades are mentioned in the recommendations. However, recommended semantic annotations of the rating scales for ACR are *Bad, Poor, Fair, Good* and *Excellent* and for DCR *Very annoying, Annoying, Slightly annoying, Perceptible but not annoying* and *Imperceptible*. In ACR stimuli are evaluated against an implicit and in DCR against an explicit reference. Other procedures of stimulus presentation in psychophysical quality tests [5], [6] as well as rating scales and combinations of these are still subject to research [7]. All these subjective test procedures deliver quality assessments for multimedia

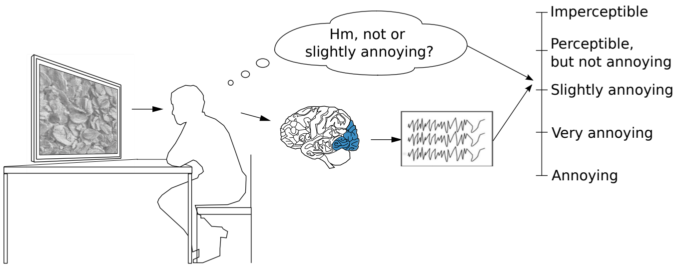


Fig. 1. From behavioral studies to BCI-based quality assessment.

signals when averaged over many subjects. These averaged assessments are then referred to as mean opinion score (MOS) [8] and considered as ground truth, e.g. for evaluating objective quality assessment algorithms.

In order to prevent observers from becoming unreliable due to fatigue, experiments should not take longer than 30 minutes [2]. This restricts the number of possible conditions that can be evaluated in one session and might make many sessions necessary. The effort of psychophysical tests is further increased by the need for having sufficient participants (15 is recommended in [2], 24 in [9]).

This motivated researchers recently to delegate the assessment of perceived multimedia quality to crowd-sourcing approaches. Here, the stimuli are brought over the internet to the participants and evaluated browser-based at their home PC in a non-lab environment. By this, subjective tests can be massively parallelized [10].

However, crowdsourced approaches and lab-based approaches share the drawback that ratings are highly variable across subjects [9]; even the same subject is unlikely to give the same rating to the same stimuli if asked repeatedly [11]. Thus, quality ratings collected in psychophysical tests suffer strongly from label noise. Further, an individual rating is the result of a conscious process and by this is prone to subjective factors, such as decision strategies or expectations [11]. Contextual biases may be introduced by the experimental design, where the stimulus range or the stimulus density can have an influence on the subject's rating [12]. In general, psychological scales do not conform to the laws of fundamental measurements known from natural sciences. Semantic annotation of rating scales given to participants during psychophysical tests may also fail at reflecting the participant's appraisal of the stimulus and by that mislead the subject's response. Another limitation of psychophysical approaches to multimedia quality assessment is the restriction to supra-threshold stimuli and the insensitivity to sub-threshold stimuli. This is crucial when it comes to short-term assessment of non-instantaneously perceivable phenomena such as visual fatigue or nausea [13].

## III. Brain Computer Interfaces for Multimedia Quality Assessment

Today, electroencephalography (EEG) is one of the most popular methods used for the acquisition of neural data in BCI.

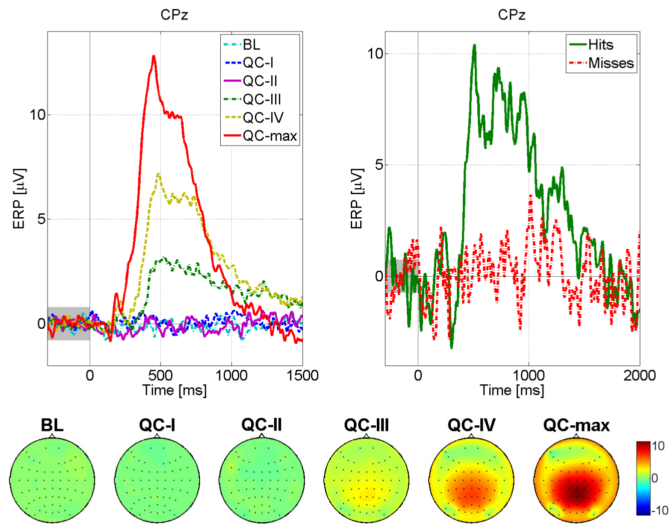


Fig. 2. From an ERP study on video quality [16]: Grand average ERP plots for different distortion levels. Top left: ERP for undistorted trials and the different quality changes at channel CPz (distortion magnitude is increasing from BL to QC-max). Top right: ERP for an intermediate distortion level for one subject, subdivided in trials, where no distortion was reported (misses) and trials, where a distortion was reported (hits). Bottom: Scalp topographies for all channels. Each circle depicts a top view of the head, with the noise pointing upwards. Colors code the mean voltage for the time interval from 400–700 ms after the introduction of a distortion into the video signal.

Since its first recording in 1924 by Hans Berger, EEG also has a long history in cognitive psychology and cognitive science, where it is used to study neural processing of sensory stimulation (e.g. [14]). Here, neurophysiological approaches are complementary to classical psychophysical ones (in fact, Gustav Theodor Fechner already 1907 postulated *inner psychophysics* as a neural foundation of *outer psychophysics* [15]). However, as outlined in Sec. II, the conventional approach to multimedia QA follows the line of (outer) psychophysics. Recently, multimedia engineers and researchers started to shift from behavioral psychophysical experiments to studies addressing inner psychophysics by adopting methods known from BCI and machine learning to assess perceived quality of multimedia signals or systems (Fig. 1).

### A. Audio Signals

Audio signals, such as speech signals, where among the first modalities studied in the context of EEG-based quality assessment. In [18], event-related potentials (ERPs) are used as a quantitative measure for perceived quality of audio signals. As stimulus a phoneme /a/ at varying quality levels was presented to the subjects for 160 ms. It is shown that the latency of the P300 is decreasing, while the amplitude is increasing with decreasing quality of the stimulus in an oddball paradigm. By applying a classifier based on shrinkage linear discriminant analysis (LDA) [19], [20], distortions below the threshold of conscious perception are detected for 2 of 11 subjects. This work is extended in [21], where longer and more realistic auditory stimuli of lengths of 200, 1200 and 8000 ms (phoneme, word, sentence) distorted by a real-world

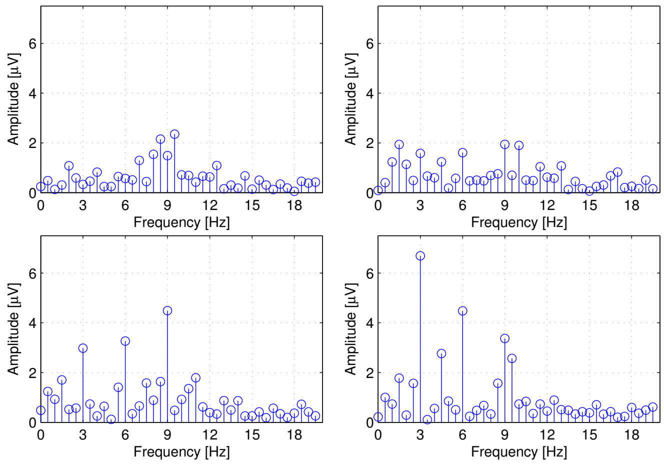


Fig. 3. From a SSVEP study on image quality assessment using a stimulation frequency of 1.5 Hz [17]: Spectrum of EEG signal at Oz for increasing distortion from top left to bottom right. On a dataset of 6 different source images at 6 different distortion levels each a correlation between MOS and amplitude at 6 Hz of $|r| = 0.93$ is reported.

transmission system were presented and similar effects as for phoneme length stimuli are found. Due to the low signal-to-noise ratio (SNR) of EEG signals, commonly a lot of trials have to be collected. The authors of [22] explore single trial methods for the analysis of neural correlates of speech quality and the detection of perceived distortions. It is shown that for auditory stimuli components reflecting the perception of distortions are assessable in early components already. Further, by combining behavioral and neural data, the sensitivity of the experiment is significantly increased.

The influence of low-quality audio signals on fatigue is studied in [23], where $\alpha$- and $\theta$-activity is shown to be increased for audio signals subject to distortions introduced by bandwidth limitations.

## B. Visual Signals

*1) 2D visual signals:* The brain response to JPEG compression artifacts in images is studied in [24] based on an oddball paradigm. As shown for speech signals, the elicited ERP component is reduced in latency and increased in amplitude for increasing visual quality. In follow-up studies, similar results are reported for different kinds of artifacts in video clips [25]. Fig. 2 shows typical ERP-responses and related scalp topographies. Using principal component analysis (PCA) for dimensionality reduction and support vector machine (SVM), a classification accuracy of 76.5% for the most obvious and of 73.5% for the less obvious distortions is reported in [26] for trial correctly classified behaviorally. For different types of distortion, mean single-trials classification accuracy of up to 85% for distorted vs. undistorted images is achieved in [27] using a wavelet-based approach. In [16], recorded EEG data is filtered by an LDA filter [19], [20]. Weights for the filter are obtained based on the signed biserial correlation coefficients between trials with highest distortion and no distortion. For distortion magnitudes above the behavioral perception threshold, an area under the ROC curve (AUC) close to 1 is obtained. Although the classifications accuracies from [16], [27], [26] refer only to trials correctly classified behaviorally, for three subjects [16] also reports an average classifications accuracy of 65% for trials with a quality degradation below the behavioral perception threshold. As in [18] for speech signals, this suggests the potential of EEG to assess also subconscious processing of distortions in multimedia signals.

Similar results are reported in [28] for degradations introduced as changes in color saturation and changes in maximum luminance values of images.

It is known from other BCI-applications that systems using the steady-state visual evoked potential (SSVEP) have advantages over ERP-based approaches. Given the broad-band characteristic background noise in EEG and the narrow-band characteristic of the SSVEP, SSVEPs achieve a high SNR compared to ERPs [29]. Thus, in [30] image quality of compressed images is assessed using an SSVEP-approach at a stimulation frequency of 1.5 Hz. Using common spatial pattern (CSP) for feature extraction on the EEG-signals filtered around the 2nd and 4th harmonic (3 Hz and 6Hz) and applying LDA for classification, mean accuracies of 84% are obtained for the highest distortion level. It is also shown that the $\alpha$-activity predicts (Pearson correlation $r \approx -0.64$) participant-wise classification accuracy. Based on the same EEG-recording, in [17], [31] a significant correlation of $r = -0.93$ between the behaviorally assessed MOS values and the amplitude at the 4th harmonic (6 Hz) of the stimulation frequency at the Oz-electrode is shown. The increase of the spectral amplitudes on the harmonic frequencies with increasing level of distortion is shown in Fig. 3.

*2) 3D visual signals:* In stereoscopic displays, depth impression is created by presenting two images, captured or rendered from a slightly offset camera positions, separately to the left and right eye. Displays using shutter glasses trade in temporal resolution for depth impression, as the two images are alternated on screen while the shutter glasses open and close correspondingly. Polarization-based systems present frames on screen polarized pixelwise alternating in different circular directions. Light polarized in one or the other directions passes the according glass and enters the corresponding eye. By this, depth impression is traded in for luminance. Additional to the distortions known from 2D visual signals, the perceived quality of 3D visual signals can be affected by crosstalk, misalignment of the stereo image pair (e.g. vertical disparities) or the vergence-accommodation conflict [13]. The neural workload imposed to the viewer may result in visual discomfort or fatigue that might not become conscious within an instant [13]. [32] evaluates the relation between the power in different oscillatory neural bands and MOS values for 3D videos subject to compression artifacts. In a single channel analysis, correlations of $|r| \approx 0.25$ are reported. Exploratory studies on the visual discomfort show a relation to changes in band power [33], [34] and ERP [35], [36]. However, all of these studies compare to the presentation of 2D signals. [37] addresses the neural classification of comfort zones in 3D viewing positions. Using ERPs and regularized Eigen Fisher spatial filters for feature extraction and shrinkage LDA for classification, a mean classification accuracy of 63.3% is reported. The influence of the shutter frequency of shutter glasses on the neural workload of the viewer is evaluated in [38]. Shutter frequencies from 39 Hz to 97 Hz are used to present stereoscopic stimuli. Neural correlates of the flicker introduced by the opening and closing of the shutter glasses could be identified up to a frequency of 67.2 Hz, well above the behaviorally estimated flicker fusion threshold at 47.4 Hz. It is concluded that the risk of reduction in quality of experience and usability can be reduced by using higher frequencies for shutter glasses.

## IV. Discussion and Challenges

As the previous section has shown, some research has been done already on BCI-based assessment of perceived quality of multimedia signals and systems. However, most of the work covers more exploratory studies and presents proofs of concept.

Little is known yet about the properties and limits of the identified paradigms. This lack of knowledge suggests a lot of potential regarding the optimization of the experimental setups. E.g. in [22], it is shown that for the neural assessment of perceived speech quality the early auditory components are enough to classify distorted from undistorted audio signals. This indicates that EEG recordings could be shortened at least for stronger distortions. Another example addresses experiments exploiting SSVEP for quality assessment; it can be shown (e.g. in [39] for perceptual threshold estimation or in [40] for classification in BCI) that the stimulation frequency in SSVEP recordings has an influence on the SNR of the recorded signal and by that on the system performance. The optimal stimulation frequency for perceived quality assessment is yet unknown. Fig. 4 shows preliminary results from a study on the relation between stimulation frequency and SNR of SSVEP for image quality assessment.

Besides the experimental optimization of identified paradigms, more challenges lie in the evaluation, development and optimization of algorithms for analyzing EEG-data recorded in quality assessment studies. LDA-based filters [19], [20] and different variants of CSP [19], [20], [41], [42],

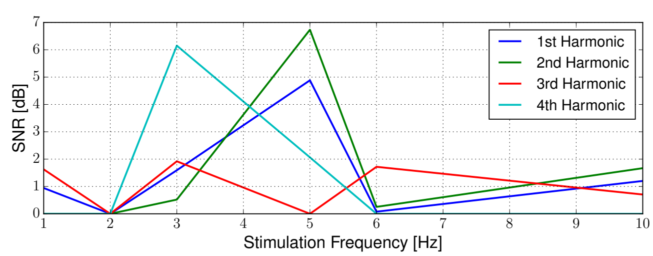


Fig. 4. SNR vs. stimulation frequency estimated for the first four harmonics of the stimulation frequency for one subject in an image quality assessment study using SSVEP. For the 1st, 2nd and 4th harmonics clear peaks are visible.

[43] are used for feature extraction and shrinkage LDA for classification [16], [30], [22]. However, it is not clear that these algorithms are optimal for the domain of quality assessment, as above the perception threshold, perceived quality can be assumed to be a quasi-continuous value, as inherently reflected by MOS values. Interesting approaches for the optimization of spatial filters are trial-to-trial or subject-to-subject correlation [44], [45] as well as recently proposed interpretable deep neural networks [46], and multimodal [47] and multisubject [48] approaches that combine multiple sources of information and thus may perform better than classical methods.

While SSVEPs allow for a faster and more reliable recording of stimulus related EEG signals than ERPs [29], a final conclusion regarding the optimal type of evoked potential for QA can not be drawn as no comparative evaluation of different approaches is available. Part of this problem is that no feasible database of test material for psychophysiological quality assessment has been adopted yet.

For the evaluation of BCI-based approaches to QA a set of criteria can be defined in order to allow for the comparison of these methods:

- How many trials do we need per subject: Condition-wise variance per subject
- How many subjects do we need: Condition-wise variance over all subjects
- Which kind of distortions can we differentiate: Granularity regarding neighbored distortion levels

Ultimately, for this, new QA methods have to be compared to the conventional, psychophysical ones. Thus, whenever a BCI-based method is evaluated beyond proof of concept, either classification accuracies or correlations are reported. Usually, either the distortion in terms of objective signal difference (such as peak signal-to-noise ratio (PSNR)), distortion parameter (such as quantization step size) or the behavioral response is considered as ground truth. Within the context of assessment of perceived quality, there is probably not much sense in taking some notion of distortion parameter into account as a reference as we want to learn *if* a specific signal or *which* of several signals is perceived distorted. Typically, this is not strongly correlated with distortion parameters over different types of contents. However, considering the behavioral responses, e.g. MOS values, as ground truth should allow us to compare thresholds and study subconcious processes. Further, as discussed in Sec. II, behavioral responses are outcome of a noisy process itself [49]. This also has to be considered in the evaluation and comparison of BCI-based quality assessment methods.

In psychophysical experiments, subjects are not only prone to biases, but some subjects might also be generally unreliable [11]. Thus, participants in behavioral studies are screened after the rating experiment in order to identify these unreliable subjects and remove their data from the analysis [2] (note that the screening process is no *absolute* process, but only *relative* to the whole set of subjects). For BCI-based approaches, $\alpha$-activity is shown to be related to subjects reliability [30], but more precise predictors have to be developed for quality assessment as done for BCI [50], [51].

While BCI-based assessment of perceived quality are potentially able to refine psychophysical methods (as in [22] for audio and [16] for video), offer more insights into distortion sensitivity close to the perception threshold, and teach researchers about sensory and cognitive processes underlying experience of media quality, they also might change the way of how quality of 3D media is assessed. Visual fatigue, visual discomfort and nausea are still very ill-defined concepts in the field of quality assessment [13] and they are usually not perceivable instantaneously by viewers, but become noticeable only after some time [13]. This renders the psychophysical assessment of 3D visual media quality difficult and time consuming. Here, BCI-based assessment has the potential to reduce the duration of system evaluation tremendously, as quality related neural correlates can be recorded before subjects notice degradations consciously. This would not only have influence on how perceived quality is assessed for entertainment applications, but also have strong impact on emerging applications such as telemedicine, as those applications usually work based on stereoscopic visualization techniques.

This becomes even more manifest when we move from stereoscopic media to VR, as VR adds the notion of *immersiveness* or *presence* to the features *naturalness* (as in 2D or audio), fatigue, discomfort, and nausea (as in 3D media): Obviously, every experience of immersiveness would be gone if subjects were asked about its quality. BCI-based QA can play its real strength here as it can deliver information about perceived quality in real-time without affecting the focus of attention and by that the feeling of immersiveness or presence of the user.

In this paper we explained why multimedia quality assessment is an important field in modern information technology. We gave an overview over conventional approaches and their drawbacks, showed how these shortcomings motivate the use of BCI for quality assessment and outlined recent advances in BCI-based quality assessment. We concluded with a discussion of current challenges in the field of BCI-based quality assessment. By this we hope to have demonstrated that multimedia quality assessment is an interesting and diverse field of research and application for BCI with a lot of open challenges ranging from rather fundamental questions regarding the nature of neural correlates of experience, over data processing techniques and their optimization to the design of quality assessment paradigms.

## References


[1] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz, "Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 82–90, 2008.

[2] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep., International Telecommunication Union, Geneva, Switzerland, 2012.

[3] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Tech. Rep., International Telecommunication Union, Geneva, Switzerland, apr 2008.

[4] ITU-T Recommendation P.800, "Methods for objective and subjective assessment of quality," Tech. Rep. ITU-T Rec. P.800, International Telecommunication Union, 1996.

[5] J.-L. Blin, "SAMVIQ–Subjective assessment methodology for video quality," *Rapp. Tech. BPN*, vol. 56, 2003.

[6] D. M. Rouse, R. Pépion, P. Le Callet, and S. Hemami, "Tradeoffs in subjective testing methods for image and video quality assessment," in *IS&T/SPIE Electron. Imaging*. International Society for Optics and Photonics, 2010, pp. 75270F—-75270F.

[7] P. Garcia Freitas, J. A. Redi, M. C. Q. Farias, and A. F. Silva, "Video quality ruler: A new experimental methodology for assessing video quality," in *Qual. Multimed. Exp. (QoMEX), 2015 Seventh Int. Work.* IEEE, 2015, pp. 1–6.

[8] ITU-T Recommendation P.800.1, "Mean Opinion Score (MOS) terminology," Tech. Rep., International Telecommunication Union, 2006.

[9] M. H. Pinson, L. Janowski, R. Pepion, Q. Huynh-Thu, C. Schmidmer, P. Corriveau, A. Younkin, P. Le Callet, M. Barkowsky, and W. Ingram, "The influence of subjects and environment on audiovisual subjective tests: An international study," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 6, pp. 640–651, 2012.

[10] S. Bosse, M. Siekmann, J. Rasch, T. Wiegand, and W. Samek, "Quality Assessment of Image Patches Distorted by Image Compression Using Crowdsourcing," in *Proc. IEEE Int. Conf. Multimed. Expo*, 2016.

[11] L. Janowski and M. Pinson, "The Accuracy of Subjects in a Quality Experiment: A Theoretical Subject Model," *IEEE Trans. Multimed.*, vol. 17, no. 12, pp. 2210–2224, 2015.

[12] D. R. Riskey, "Use and Abuses of Category Scales in Sensory Measurement," *J. Sens. Stud.*, vol. 1, no. 3-4, pp. 217–236, dec 1986.

[13] M. Urvoy, M. Barkowsky, and P. Le Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Ann. des Telecommun. Telecommun.*, vol. 68, no. 11-12, pp. 641–655, 2013.

[14] S J Luck, "An Introduction to the Event-Reated Potencial Technique," p. 376, 2005.

[15] G. T. Fechner, *Elemente der Psychophysik. 2*, Breitkopf & Härtel, 1907.

[16] 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.

[17] 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.* oct 2014, pp. 1987—-1991, IEEE.

[18] A. K. Porbadnigk, J.-N. Antons, B. Blankertz, M. S. Treder, R. Schleicher, S. Möller, and Gabriel Curio, "Using ERPs for assessing the (sub) conscious perception of noise," in *Eng. Med. Biol. Soc. (EMBC), 2010 Annu. Int. Conf. IEEE*. IEEE, jan 2010, vol. 2010, pp. 2690–2693.

[19] 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.

[20] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE Signal Process. Mag.*, vol. XX, pp. 1–12, 2008.

[21] 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.

[22] 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, pp. 056003, oct 2013.

[23] J.-N. Antons, R. Schleicher, S. Arndt, S. Möller, and G. Curio, "Too tired for calling? A physiological measure of fatigue caused by bandwidth limitations," *2012 4th Int. Work. Qual. Multimed. Exp. QoMEX 2012*, pp. 63–67, 2012.

[24] 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.

[25] M. Mustafa, L. Lindemann, and M. Magnor, "{EEG} Analysis and of Implicit and Human Visual and Perception," in *CHI*, 2012.

[26] 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, pp. 53, 2011.

[27] 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.

[28] Y. He, Y. Tu, L. Wang, L. Jia, J. Guo, and Y. Chen, “A Study on the Relationship Between Event-Related Potentials and Image Quality Variation,” *J. Disp. Technol.*, vol. PP, no. 99, pp. 1–10, 2016.

[29] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottereau, and B. Rossion, “The steady-state visual evoked potential in vision research : A review,” *J. Vis.*, vol. 15, no. 6, pp. 1–46, 2015.

[30] 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, pp. 026012, 2015.

[31] 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.* International Society for Optics and Photonics, 2015, vol. 9599, pp. 959914–959914, International Society for Optics and Photonics.

[32] 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.

[33] C. Chen, K. Li, Q. Wu, H. Wang, Z. Qian, and G. Sudlow, “EEG-based detection and evaluation of fatigue caused by watching 3DTV,” *Displays*, vol. 34, no. 2, pp. 81–88, 2013.

[34] C. Chen, J. Wang, K. Li, Q. Wu, H. Wang, Z. Qian, and N. Gu, “Assessment visual fatigue of watching 3DTV using EEG power spectral parameters,” *Displays*, vol. 35, no. 5, pp. 266–272, 2014.

[35] H. O. Li, J. Seo, K. Kham, and S. Lee, “Measurement of 3D Visual Fatigue Using Event-Related Potential (ERP): 3D Oddball Paradigm,” in *2008 3DTV Conf. True Vis. - Capture, Transm. Disp. 3D Video*. may 2008, pp. 213–216, Ieee.

[36] H. Cho, M.-K. Kang, K.-J. Yoon, and S. C. Jun, “Feasibility study for visual discomfort assessment on stereo images using EEG,” *2012 Int. Conf. 3D Imaging*, pp. 1–6, 2012.

[37] J. Frey, A. Appriou, F. Lotte, and M. Hachet, “Estimating visual comfort in stereoscopic displays using electroencephalography: A proof-of-concept,” in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2015, vol. 9299, pp. 354–362.

[38] 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, pp. 016003, 2016.

[39] 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.

[40] 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.,” *J. Neural Eng.*, vol. 13, no. 1, pp. 016014, 2015.

[41] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms,” *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 2, pp. 355–362, 2011.

[42] W. Samek, M. Kawanabe, and K.-R. Müller, “Divergence-based framework for common spatial patterns algorithms,” *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.

[43] M. Kawanabe, W Samek, K.-R. Müller, and C. Vidaurre, “Robust common spatial filters with a maxmin approach,” *Neural Computation*, vol. 26, no. 2, pp. 1–28, 2014.

[44] M. Gaebler, F. Biessmann, J.-P. Lamke, K.-R. Müller, H. Walter, and S. Hetzer, “Stereoscopic depth increases intersubject correlations of brain networks.,” *Neuroimage*, vol. 100, pp. 427–34, oct 2014.

[45] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, “Correlated Components of Ongoing EEG Point to Emotionally Laden Attention A Possible Marker of Engagement?,” *Front. Hum. Neurosci.*, vol. 6, no. May, pp. 112, 2012.

[46] I. Sturm, S. Bach, W. Samek, and K.-R. Müller, “Interpretable deep neural networks for single-trial eeg classification,” *arXiv*, , no. 1604.08201, 2016.

[47] S. Dähne, F. Bießmann, W. Samek, S. Haufe, D. Goltz, C. Gundlach, A. Villringer, S. Fazli, and K.-R. Müller, “Multivariate machine learning methods for fusing multimodal functional neuroimaging data,” *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1507–1530, 2015.

[48] S. Fazli, S. Dähne, W. Samek, F. Bießmann, and K.-R. Müller, “Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain-computer interfaces,” *Proceedings of the IEEE*, vol. 103, no. 6, pp. 891–906, 2015.

[49] 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.

[50] 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.

[51] 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.