DiVA
http://www.diva-portal.org

Postprint

This is the accepted version of a paper published in *IEEE Journal on Selected Topics in Signal Processing*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Engelke, U., Darcy, D., Mulliken, G., Bosse, S., Martini, M. et al. (2017)
Psychophysiology-based QoE Assessment: A Survey
IEEE Journal on Selected Topics in Signal Processing, 11(1): 6-21
https://doi.org/10.1109/JSTSP.2016.2609843

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-28754

Psychophysiology-based QoE Assessment: A Survey

Ulrich Engelke, *Senior Member, IEEE,* Daniel P. Darcy, Grant H. Mulliken, Sebastian Bosse, Maria G. Martini, *Senior Member, IEEE,* Sebastian Arndt, Jan-Niklas Antons, Kit Yan Chan, Naeem Ramzan, *Senior Member, IEEE,* and Kjell Brunnström

***Abstract*—We present a survey of psychophysiology-based assessment for Quality of Experience (QoE) in advanced multimedia technologies. We provide a classification of methods relevant to QoE and describe related psychological processes, experimental design considerations, and signal analysis techniques. We summarise multimodal techniques and discuss several important aspects of psychophysiology-based QoE assessment, including the synergies with psychophysical assessment and the need for standardised experimental design. This survey is not considered to be exhaustive but serves as a guideline for those interested to further explore this emerging field of research.**

***Keywords*—Psychophysiology, quality of experience, electroencephalography, near-infrared spectroscopy, electrocardiography, electrodermal activity, eye tracking, pupillometry.**

I. Introduction

QUALITY of Experience (QoE) has grown to a multidisciplinary research field evaluating the relationship between a wide range of human and system related factors. Understanding QoE is critical for any organisation with an interest in providing the best experience to their customers given certain system and service limitations. Methods for evaluating QoE have grown increasingly sophisticated and a substantial body of research has emerged over the years with earlier works focusing on auditory [1] and visual [2] quality perception. With the advent of advanced media technologies and the notion of optimising an overall experience, more recent work has defined QoE in the context of a multisensory experience [3], [4].

Psychophysical assessment scales and open ended questionnaires are typically used, respectively, to quantitatively and qualitatively evaluate a user's media perception. These methods depend on conscious responses and often do not provide sufficient insight into underlying perceptual and cognitive processes. Participants may also have difficulties communicating their assessment given a particular scale that does not reflect well their internal perception. These issues are aggravated as we go from low-level perceptual aspects, such as distortion perception, to high-level QoE concepts. We argue that psychophysical assessment alone may therefore not provide all necessary information, especially in the context of QoE of advanced media technologies, and that other implicit rather than explicit means are needed for QoE assessment.

In this article, we present an extensive overview of such alternative methods, based on psychophysiology. While psychophysical methods quantitatively evaluate the relationship between physical stimuli and the conscious perceptions thereof, psychophysiology is concerned with the physiological bases of perceptual and cognitive processes. As such, it measures implicit rather than explicit responses to physical stimuli and thus overcomes the problem of potentially misleading rating scales and conscious decision making. Importantly, we do not suggest replacing psychophysical methods with psychophysiological ones. Instead, we propose that concurrent deployment of both methodologies, where applicable, may lead to deeper insights into high-level QoE attributes. A wide range of physiological measurement methods exist, not all of which are suitable for studies in a QoE context. In this article we focus on methods that are most accessible and promising for meaningful exploration of multimedia QoE. Towards this end, we provide background information on psychophysiology and introduce a classification of relevant methods. We then survey a wide range of psychophysiological methods and discuss important aspects of using physiological measurements in the context of QoE. With this article, we hope to provide a valuable framework for researchers interested in exploring psychophysiological approaches for QoE assessment of advanced media technologies.

II. Psychophysiology and QoE

Psychophysiology is concerned with the measurement of physiological signals and psychological correlates thereof. Many neurophysiological methods are specifically targeted to the Central Nervous System (CNS) and have been utilised for almost a century by neuroscientists, psychologists and clinicians to build insights into the neural underpinnings of perceptual, motor and cognitive processing in the brain. While the neural code and its underlying computational mechanisms are not well understood, the features that are encoded at different levels of the sensory and motor pathways have been mapped out in reasonable detail. For instance, neuronal ensembles at higher and higher levels of the visual hierarchy are tuned to progressively larger and more abstract visual

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

UE is with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Hobart, Australia; e-mail:ulrich.engelke@csiro.au.

DD and GHM are with Dolby Laboratories, San Francisco, USA.

SB is with Fraunhofer Institute for Telecommunications, Berlin, Germany.

MM is with Kingston University, London, UK.

SA is with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, and Technische Universiät Berlin, Germany.

JNA is with Technische Universiät Berlin, Germany.

KYC is with Curtin University, Perth, Australia.

NR is with the University of the West of Scotland, Hamilton, UK.

KB is with Acreo Swedish ICT AB, Stockholm, and Mid Sweden University, Sundsvall, Sweden.

Manuscript received September 5, 2016.

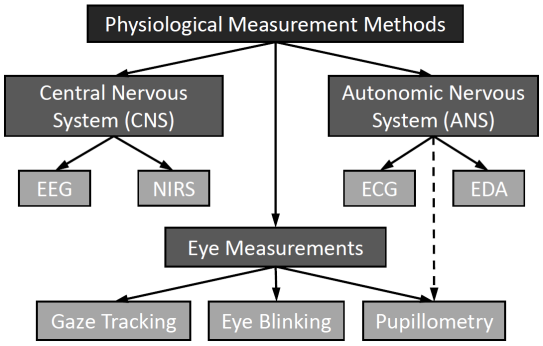


Fig. 1: Classification of physiological measurement methods.

features [5]. Moreover, the mechanisms that bind sensory information across modalities as well as the transformations that occur from sensory to motor coordinate frames are beginning to be unravelled [6] [7]. Considerable progress has also been made in characterising how cognitive processes interact with sensory representations, within and across brain areas [8]. Neural correlates of attention, decision-making, working memory, and executive control have been characterised through extensive experimental studies involving both animals and humans [9]. Many of these well-established neurophysiological signatures hold potential to be harnessed for QoE assessment in audiovisual entertainment and other technologies that interface directly with the sensory pathways, and ultimately, could enable reading out the cognitive state of an individual using quantifiable metrics. For example, selective visual attention paradigms that have been successful in establishing clear markers of visual task engagement may be a promising candidate for assessing viewer immersion using advanced imaging technologies [10] [11]. Particularly relevant to QoE-based neurophysiological investigations are electroencephalography (EEG) and near-infrared spectroscopy (NIRS), which are non-invasive techniques to measure brain activity and infer the cognitive state.

Beyond estimating the cognitive state of an individual from measurements derived from the CNS, physiological methods open a window to measure and exploit the inner workings of the peripheral nervous system. Specifically, the sympathetic division of the autonomic nervous system (ANS) and its 'fight or flight' response [12] can be quantified reliably using electrocardiography (ECG), electrodermal activity (EDA) and pupil diameter. Importantly, the CNS and ANS overlap considerably in their representations, therefore, careful experimental design is essential to dissociate modulations caused by autonomic versus cognitive phenomena. At the same time, thoughtful design of decoding algorithms whose goal is to appropriately fuse information across multimodal measurements is an important research direction.

The eyes also provide a window to the brain, a claim supported by decades of observations demonstrating a relationship between cognition and eye measurements. For instance, eye movements provide valuable insight into overt visual attention in addition to covert attention captured through EEG measurements [13]. Furthermore, eye blink rate is known to relate to visual fatigue [14] and pupil dilation to cognitive load [15], both of which are relevant to QoE assessment. Of practical relevance, the eye measurement may provide insight into cognitive activity that is not easily observable through other methods.

Given the above, we distinguish between three classes of physiological measurements in this survey, relating to the CNS, ANS, and eye measurements. Within each class, we limit ourselves to methods applicable in a QoE context and we exclude prohibitive methods that are typically seen only in clinical contexts, such as functional Magnetic Resonance Imaging (fMRI). An overview of these methods is given in Fig. 1. The treatment of each method in the article is balanced to reflect the existing research as it relates to QoE. While gaze tracking has received considerable attention in the QoE research community [13], other physiological measurements have largely been deployed in other domains. More recently, however, the QoE community has begun to explore physiology and in particular EEG as a means to assess multimedia QoE and initial outcomes appear promising.

III. Central Nervous System

A. Electroencephalography

1) Psychological and physiological processes: Electroencephalography (EEG) is a minimally invasive electrophysiological measurement of voltage changes at the scalp arising from electrical activity of neuronal ensembles that underlie cognitive states and experience [16]. Electrophysiological measurements have been studied for many decades with Hans Berger credited as the first to record EEG in humans [17] in 1924. While the precise links between EEG and psychological phenomena are incompletely understood, several relations are widely accepted. For example, event-related potentials (ERPs), large scale electrical events related to high- and low-level sensory and cognitive processing, have been thoroughly characterised. ERPs consist of stereotypic changes in electrical activity usually evoked by time-locked sensory stimuli and related cognitive events. They are characterized by their time-dependent amplitude according to a common nomenclature, with the first letter referring to the polarity of a particular component and subsequent number(s) indicating latency (in ms) or relative position in the order of components. For example, the well-known P300 component exhibits a positive peak around 300 ms after stimulus onset. The amplitude of its subcomponent P3b is known to increase with decreased expectation of a stimulus, thus indicating the novelty of a stimulus. Other ERPs have been shown to be involved with object representation and memory operations in a variety of task behaviours.

Another neurophysiological response to temporally isolated stimuli is the steady-state visually evoked potential (SSVEP) [18]. While ERPs are typically observed in response to surprising or novel stimuli, SSVEPs are observed in response to sustained, periodic stimuli. The response to this kind of stimulation is usually stable in amplitude and phase as suggested by its name. Sensory drive elicited by such periodic stimuli

TABLE I: Neural frequency bands, their predominant locations on the scalp, and related psychological processes.

Band	Frequency [Hz]	Scalp location	Psychological processes
δ	< 4	Frontal/Posterior	Present during deep sleep
θ	$4 - 8$	Various	Lack of attention, drowsiness
α	$8 - 13$	Posterior/Occipital/Temporal	Alertness decrease, fatigue
β	$13 - 31$	Frontal/Parietal	Alertness, focused attention
γ	> 31	Various	Linguistic processing, associative learning

results in increased narrowband EEG spectral power at the tagged stimulus frequency in corresponding sensory areas of the brain. SSVEPs are typically described by their amplitude, phase and spatial channel distribution (for the tagged frequency and its associated harmonic components) [18].

Neural oscillations due to synchronization of neuronal ensembles are measurable as peaks in the power spectrum of the EEG. Such rhythmic activity has been linked to a large number of psychological processes. Table I provides an overview of some of the various prominent neural rhythms found in the brain.

2) Experiment design and Data Analysis: The types of EEG responses desired will invariably guide experimental design and data analysis. For example, spectral analysis is typically used in conjuction with longer time-scale stimuli to study cognitive states that covary with power modulations in different frequency bands. Most prominently used in the context of QoE assessment are θ $(4-8Hz)$ waves, which serve as an indicator of decreased attention and increased drowsiness as well as α $(8 - 13Hz)$ waves, which indicate decreased alertness and increased relaxation or fatigue.

Both ERPs and SSVEPs are analysed in the time-domain. For ERPs, the presentation of stimuli is typically short in duration, with several trials generally averaged together. The P300 component, commonly used in the context of QoE assessment, is typically associated with a 300 ms latency from stimulus onset, but can be significantly higher and reach 500 ms or more. An example of an ERP elicited by video distortion is plotted in the left panel of Fig. 2. It can be clearly seen that the magnitude of the P300 response increases with the degree of distortion. In SSVEP experimental paradigms, stimuli are usually presented in an alternating fashion, toggling between a test and a reference condition. The amplitude of the SSVEP can be considered a neural marker of the sensory or perceptual difference of the alternating stimuli, whereas the phase (the frequency domain representation of delay) reflects the complexity of neural processing. An example of the change in SSVEP amplitude is illustrated in the right panel of Fig. 2.

The signal-to-noise ratio (SNR) of EEG signals is typically very low, making single trial quality assessment difficult. This can be overcome by trial repetition or a combined analysis of signals measured at different scalp locations using spatial filters. These filters project the recorded data from the sensor-space to a subspace determined by an optimization criterion that enhances signal recovery.

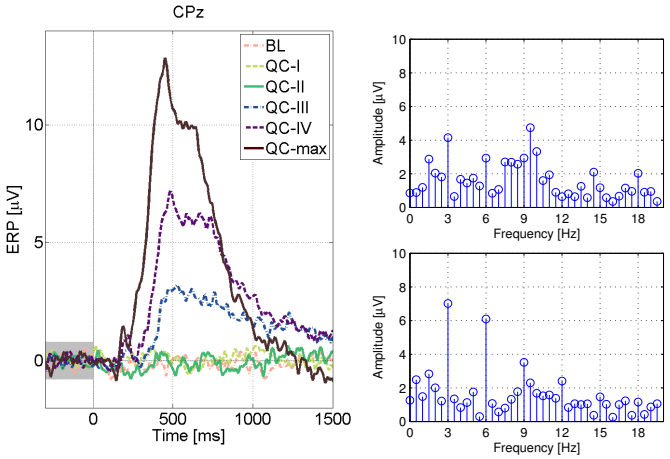


Fig. 2: Left panel: ERP measured from videos of different distortion magnitudes: undistorted (BL) and distortion increases from QC-I to QC-max [19]. Right panel: SSVEP measured from images with low (top) and high (bottom) distortions [20].

3) Prior art: One of the first studies using EEG for quality assessment was conducted in the auditory domain. Antons et al. [21] asked participants to judge whether they perceived signal-correlated noise in short (syllables) and long stimuli (words, sentences). The peak amplitude of the elicited P300 was shown to increase significantly with an increase in distortion level [22]. The latency of the P300 peak was shorter with stronger distortions. Furthermore, distorted stimuli that were rated as undistorted had a similar trend in the ERP as those trials rated as distorted. High classification rates for these trials could be obtained and it was conjectured that these degradations are processed unconsciously [23]. In related studies, a decrease in speech stimulus quality was associated with an increase in sleepiness. This effect was measured by changes in the α band in response to speech stimuli of long duration (audio books) in constant [24] and varying quality conditions [25]. Arndt et al. [26] studied the QoE of text-to-speech systems. Strong correlations were found between the P300 amplitude and speech intelligibility and quality. Laghari et al. [27] further showed that the quality of text-to-speech is correlated with asymmetric activation in the frontal head region and activity in orbito-frontal cortex (OFC).

Limited work has been done on imaging applications. Bosse et al. [28] showed that the 2^{nd} and 4^{th} harmonic components of SSVEP responses, elicited by a stimulation of 1.5Hz, are correlated with image quality. Spatial distributions of activity were in line with neurophysiological interpretations of sensory processing. Acqualagna et al. [29] classified the magnitude of perceived image distortions based on common spatial patterns (CSP) of various frequency channels.

Similar approaches to those in speech quality have been used in video and audiovisual quality assessment [30]. In a series of experiments using spoken syllables, it was shown that the evoked P300 increased with the distortion level of

audiovisual sequences [31]. For longer sequences, low-quality conditions led to higher α and θ band power [32]. Scholler et al. [19] recorded EEG data at different scalp positions, which was decoded using linear discriminant analysis to detect the perception of video distortion. Donley et al. [33] studied the impact of various levels of synchrony of wind, vibration and light on audio-visual sequences. The temporal and occipital lobes were found to have more activity during asynchronous and synchronous effects, respectively. Arndt et al. [34] investigated the effect of audio and video quality using EEG and eye tracking parameters. This study concluded that α activity was significantly predictive of video quality. Moon et al. [35] investigated the perception of high dynamic range (HDR) and low dynamic range (LDR) videos and showed a significant power difference in the γ band.

Beyer et al. [36] varied compression rates of video for cloud games and found a strong correlation between EEG and cloud gaming QoE. Kroupi et al. [37] analysed degradation of 2D and 3D videos for seven one-minute long recording sequences from a music festival. Each sequence was presented in a high and low quality version using 2D and 3D rendering. Spectral power in the EEG was correlated with self-reported quality judgements. EEG recordings showed high frontal asymmetry in the α band, which reflected emotional affect towards the two different quality levels.

Chen et al. [38] investigated visual fatigue for 2DTV and 3DTV viewing using 16-channel EEG measurements. Significant decreases in gravity frequency and power spectral entropy, related to alertness level decline, were observed in several brain regions after extended 3DTV viewing. Based on these findings and psychophysical responses, an accurate evaluation model for 3DTV fatigue was established. A related study [39] compared the 2D/3D changes of energy values in four spectral bands $(\alpha, \beta, \gamma, \delta)$ with four fatigue-related parameters $(\alpha+\delta)/\beta, \alpha/\beta, (\alpha+\delta)/(\alpha+\beta)$ and δ/β. All bands except the θ rhythm changed significantly when subjects viewed 3DTV. In particular, the energy decreased in α and β frequency bands while δ activity increased significantly. All four fatigue related parameters showed a significant increase, with the highest increase for $(\alpha + \theta)/\beta$.

More comprehensive reviews on using EEG in QoE research can be found in [40], [41].

B. Near-Infrared Spectroscopy

1) Psychological and physiological processes: near-infrared spectroscopy (NIRS) or functional NIRS (fNIRS) is a neuroimaging technology that uses hemodynamic measurements to infer brain activity. In comparison to EEG, NIRS is a more indirect method that measures oxygen concentration of the blood through light emitting and detecting diodes. Levels of oxygenated ($\Delta[HbO]$) and deoxygenated ($\Delta[HbR]$) blood have different absorption spectra. This allows using wavelength differences of reflected light (typically between 650-950 nm) to form inferences about activated and non-activated areas in the brain. While the temporal resolution of EEG is very high, the blood flow changes measured by NIRS occur with latencies of several seconds. On the other hand, NIRS provides a much better spatial resolution than EEG and sources of activation can be approximated more precisely.

2) Experiment design and analysis: The setup of a NIRS experiment is largely similar to EEG with some important differences to consider. The significant latency of blood flow change is important to take into account when designing an experiment using NIRS measurements. In order to avoid interfering with previous stimuli, an inter-stimulus interval of at least 10 s should be used to allow for the blood flow to normalise again after stimulus presentation. As a result, only a limited number of stimuli can be used in order to keep the experiment duration reasonable for participants. Alleviating this constraint to some degree is the fact that NIRS sessions typically do not require as high a number of trial repetitions when compared to EEG experiments.

Light emitting diodes need to be placed on the participant's head using a cap similar to EEG. The installation of the diodes, however, requires careful effort to remove hair at each sensor location to eliminate interfering reflection. This is a strong limitation of this method as participants may not agree to have even small portions of their hair shaved.

NIRS signals are analysed to extract the peak time and peak amplitude of the $\Delta[HbO]$ signal, along with their rise and recovery times. $\Delta[HbO]$ peak and $\Delta[HbR]$ valley are often correlated with the blood oxygenated level dependent (BOLD) response measured using fMRI, a method which is typically prohibitive in the context of QoE.

3) Prior art: Likely as a result of the above limitations, only very limited research has been performed in the area of QoE using NIRS. A study performed by Gupta et al. [42] used NIRS while characterising different samples of synthesised speech. It was shown that the OFC exhibits higher activation for better quality synthesized speech stimuli. Du et al. [43] used NIRS measurements to show that a decrease in quality of moving images is related to an increase in frontal brain activity.

IV. Autonomic Nervous System

A. Electrocardiography

1) Psychological and physiological processes: Electrocardiography (ECG) is a time-varying measure reflecting the ionic current flow produced by the cardiac fibres contracting and relaxing with each heartbeat cycle. A single normal period of the ECG can be approximately associated with the peaks and troughs of the canonical ECG waveform. A wide range of variables are often extracted from ECG measurements with heart rate (HR) and heart rate variability (HRV) being among the most common. In the context of multimedia QoE, these are thought to relate to excitement, fatigue and discomfort [44].

2) Experiment design and analysis: The standard ECG signal is acquired via skin electrodes that are fit by three points known as Einthoven's Triangle and connected to an ECG recorder or transmitter, which measures the potential difference between pairs of electrodes in millivolts (mV). The measured HRV reflects the sympathetic/parasympathetic modulation of the time interval between two consecutive heart beats, i.e., the variation of R-R intervals, in beats per minute (bpm). Since the HRV is a time-series of non-uniform R-R intervals it

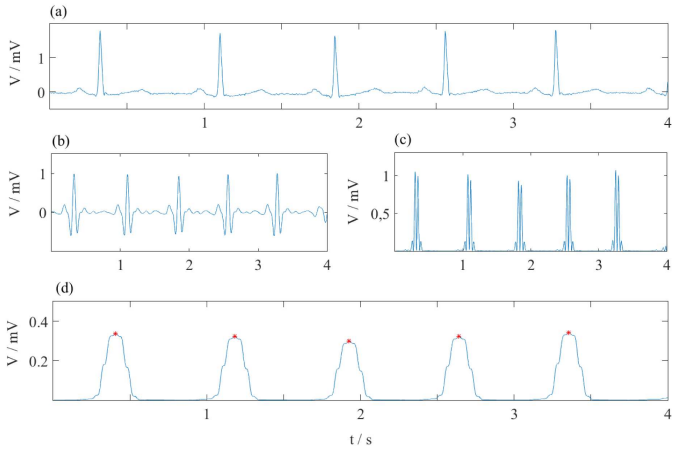


Fig. 3: Stages of the Pan-Tompkins-Algorithm [45] illustrated on a 4 s ECG signal: (a) Zero mean normalisation, (b) bandpass filtering, (c) selective differentiation and squaring (normalisation to V_{max}), and (d) moving average filtering with R-peak detection.

can be estimated by measuring the interval between two QRS complexes, and the HR can be estimated as an inverse function of the R-R interval.

Pan and Tompkins [45] proposed a real-time algorithm to measure HRV that can be used for QoE assessment [37]. It can detect QRS complexes with an accuracy higher than 99%. The stages of the algorithm are illustrated in Fig. 3. After zero-mean normalisation, the ECG signal is bandpass filtered in the range from 5 Hz to 11 Hz to attenuate noise and artefacts. The filtered signal is squared to further emphasize the QRS complex. Next, moving-window integration is performed to extract waveform information as the QRS complex corresponds to the rising edge of the integrated waveform. Finally, thresholds are adjusted within a loop and detected peaks are interpreted to determine whether it is an R-peak using decision rules. The following features are computed based on detected adjacent R-R intervals: difference between consecutive R-R intervals, power spectral density (PSD) for low frequency (LF) and high frequency (HF), the ratio of LF to HF, as well as the total power. These features can be used for classification of emotional states.

3) Prior art: ECG measurements have recently been investigated in the context of multimedia QoE assessment, often in conjunction with neurophysiological measures like EEG. Kroupi et al. [44] investigated perceived QoE for 2D and 3D multimedia stimuli, and specifically a 'Sensation of Reality', based on ECG, EEG, and respiration rate. ECG and EEG were found to be predictive of this high level QoE phenomenon, with EEG being more accurate. Barreda-Angeles et al. [46] studied visual discomfort for stereoscopic 3D on viewers' emotional states. While self-reporting did not reveal any effects of visual discomfort, physiological measures in terms of HR measured through ECG, EDA and facial electromyography (EMG) were found to strongly correlate with visual discomfort.

B. Electrodermal Activity

1) Psychological and physiological processes: Electrodermal activity (EDA) or galvanic skin responses (GSR) refer to measures of resistance on the surface of the human skin. EDA varies with the activity of sweat glands. Sweat production is driven by ANS activity and is usually elevated during arousal. In the context of QoE, a higher level of arousal is typically observed if the viewer has an enhanced level of immersion or engagement and as such it is argued to be highly dependent on the presented content and its quality.

2) Experiment design and analysis: Electrodes to measure EDA are usually put on fingers, or sometimes arms or legs. Recorded EDA measurements are usually bandpass-filtered and the amplitude over the stimulus duration is analysed. Great care must be taken to maintain a constant level of room temperature, as temperature fluctuations may also contribute to changes in EDA. Humidity and skin moisture should also not be too low in order to improve EDA signal quality. The location of sensor placement (finger, arm, leg) needs to be carefully considered with regard to the task that the participants are asked to perform. Short-term responses are typically measured with EDA while longer term changes may be attributed to external factors (e.g. sweating due to room temperature and changes in skin resistance over time). Careful temporal alignment with the stimulus is therefore essential. Stimulus repetition is often needed to obtain a clear EDA response.

3) Prior art: Lasalle et al. [47] measured EDA during viewing of audiovisual stimuli of varying quality. They did not identify a correlation of skin conductance with self-report quality responses. In an experiment using longer stimuli (15 min), Arndt et al. [32] also could not identify a direct relationship of EDA measurements with QoE. In both experiments spatial degradations were used. Wilson et al. [48] investigated EDA measures when varying the frame rate of a video sequence and were able to show an increase in EDA for lower frame rates, which can be argued to be an effect caused by induced stress of the participant. Given the limited evidence above, we argue that EDA as a physiological measurement for assessment of quality degradations needs further investigation. We also conjecture that contexts more likely to have an impact on arousal, such as immersive environments and interactive systems, may benefit more from EDA measurements compared to conventional viewing environments.

V. Eye Measurements

A. Gaze Tracking

1) Psychological and physiological processes: At any instant in time, the human eye is exposed to an abundant amount of visual information. Attentional mechanisms in the human visual system are fundamental to reducing the complexity of scene analysis. We distinguish between bottom-up and top-down attention. Bottom-up attention is reflexive, signal driven, and independent of a particular task. Top-down attention is driven by higher level cognitive factors and external influences such as contextual effects, viewing task, and personal experience. Both attentional mechanisms guide eye movements to the

most relevant information in a given context. Gaze tracking (GT) is deployed to capture these eye movements and the underlying attentional mechanisms. We note that GT records only overt visual attention (shifting of the eyes to a stimulus) but not covert visual attention (mental shift of attention), which can be measured using EEG.

2) Experiment design and analysis: Eye trackers are integrated into experiments to capture overt visual attention during visual stimulus observation. Modern eye trackers are non-invasive video based systems with an infra-red camera, typically installed under a stimulus screen, that measure corneal reflections to determine the direction of gaze. Some eye trackers come with a head rest and are installed at the location of the observer. In real-world experiments, head mounted eye trackers are used to record gaze behaviour in 3-dimensional space. In any case, the eye tracker has to be carefully time-synchronised with the stimulus presentation to accurately capture eye movements. This is particularly important if accuracy at the speed of saccadic eye movements is needed. To assure spatial accuracy, eye trackers have to be carefully calibrated to each individual observer using pre-defined calibration patterns. For long experiments, the calibration often needs to be repeated to maintain high spatial accuracy. Spatial accuracy usually diminishes towards the periphery of the visual field for which reason it is important to carefully take into account the visual angle of the presented stimulus.

Eye movements are recorded at frequencies starting typically from 50 Hz and going beyond 1000 Hz for high-end devices. The resulting raw gaze patterns are usually rather inconclusive and therefore need post-processing into more meaningful fixation density maps (FDM) [49] as shown in Fig. 4 superimposed on the image content. Typical analysis of eye tracking data includes determining fixation location, duration, and order. In the context of QoE it is critical to also compare two FDM, for instance, between distorted and reference stimuli. Some of the most commonly used comparison metrics include the Kullback-Leibler divergence (KLD), area under the ROC curve (AUC), normalised scanpath saliency (NSS), and Pearson linear correlation coefficient (PLCC). These metrics capture different aspects of the eye tracking data and are often used conjointly.

3) Prior art: Gaze tracking has been extensively investigated in the context of visual QoE assessment, the rationale being that eye gaze is directed not only by the natural content but also by potential induced distortions. The relative interplay of these and the resulting gaze behaviour is conjectured to have a major impact on the overall QoE. While the potential benefit of recording gaze behaviour has been widely recognised by the research community, the conclusions drawn from individual studies vary considerably. Covering the entire spectrum of the prior art exceeds the scope of this article and the interested reader is referred to recent survey articles [13], [50]. We discuss some relevant contributions in the following passages.

Several studies have investigated the relationship between overt visual attention in distorted and undistorted visual stimuli and the related impact on overall quality perception. In Engelke et al. [51], the impact of content saliency relative to distortion location was investigated for H.264 coded video with localised packet loss distortions. It was shown that distortions located in salient regions have a significantly higher impact on quality perception as compared to distortions in non-salient regions. This is partly attributed to a shift of attention needed in case of non-salient region distortions whereas distortions in salient regions are perceived more severely, as illustrated in Fig. 4.

Fig. 4: Shift of overt attention through visual distortions. The left frame is undistorted, middle frame is distorted in the salient region, and right frame is distorted in a non-salient region.

Global compression distortions were found by Le Meur et al. [52] to not alter viewing patterns considerably. Liu et al. [53] showed that differences in observers' viewing behaviour depends strongly on the image content. The more distinct a salient region, the higher the agreement between observers, and the larger the performance gain when including eye tracking data into objective quality metrics. Importantly, Lee et al. [54] reported that audio is a strong attractor of visual attention, and should therefore not be disregarded, but its interplay with the visual stimulus should be accounted for.

Eye tracking data is often integrated into image and video quality metrics to further improve their quality prediction performance. Ninassi et al. [55] integrated FDMs into several quality metrics to predict JPEG and JPEG2000 distortions, but no improvements were found for the considered metrics. Based on the FDMs and positive outcomes in [51], Engelke et al. [56] integrated spatial saliency weighting into a video quality metric, TetraVQM [57], successfully improving the metric's prediction accuracy. The task given to the observer is also known to influence gaze behaviour and it is generally agreed that the integration of task-free eye tracking data into quality prediction models is more successful than when using eye tracking data obtained during quality assessment task [58], [59]. While the value of task-free eye tracking data is widely accepted and several databases have been made publicly available, Engelke et al. [60] recently compared three databases and found that differences amongst their FDMs were small with low impact on the integration into quality metrics.

In summary, the value of eye tracking data for studying visual QoE is widely agreed upon and integration into computational models was generally found to be more successful for (i) video rather than image applications, (ii) local rather than global distortions, and (iii) task-free rather than task-driven eye tracking data. Despite this field being well explored, further research is needed to fully understand the interaction between overt visual attention to audiovisual content and induced distortions.

B. Pupillometry

1) Psychological and physiological processes: Tasks involving cognitive processing cause short-onset latency (100-200 ms) transient pupil dilation that peaks after a few hundred milliseconds and then rapidly reverts following task completion. Pupil responses have been shown for cognitive load and reasoning, memory, visual attention, and language processing.

In constant low-light conditions, pupil dilation has long been known to involve the activation of the locus coeruleus (LC) and its neuromodulatory influence through norepinephrine (NE) signaling pathways. NE released from the LC activates alpha-adrenoreceptors of the iris dilation muscles as well as postsynaptic alpha2-adrenoreceptors of the Edinger-Westphal nucleus, which innervates the ciliary ganglia. More recently it was shown using extracellular recordings and stimulation in LC with simultaneous monitoring of pupil diameter fluctuations that there is a causal functional relationship [61]–[63].

The LC makes widespread projections and functional interactions with many brain regions from cortex and cerebellum to other brainstem structures and spinal cord. It contributes ascending pathways responsible for arousal and engagement with novel environmental stimuli, e.g., optimizing the levels of exploration versus exploitation in adaptive behaviour, and is implicated in cognition, emotional processing, perceptual rivalry and memory retrieval. LC is not the exclusive site of activity coupled to the pupil. For instance, the pupillary light-reflex via the iris sphincter muscle is served via distinct parasympathetic pathways, and connections between the eye and other modulating brain structures are not completely understood. However, it is known that both sympathetic and parasympathetic nervous system activation affects the radial dilator muscles of the pupil and muscles in the iris, leading to pupil dilation as sympathetic activation increases and constriction with activation decrease and an opposing response for parasympathetic activity [64] [65].

2) Experiment design and analysis: Key parameters for collecting pupillometry data are speed, resolution and accuracy of eye tracking. Sufficient temporal resolution and precision are required to track the time course of change on the order of milliseconds and synchronise with external task-related stimuli. Eye tracking systems vary from tens of hertz to several kilohertz for high-speed systems used to acquire other events like microsaccades, small involuntary eye movements that occur during fixation. Increased image resolution and repeated trials averaging enables sensitive tracking of pupillary size changes up to hundredths of a millimeter.

Image processing algorithms can be used to reliably track pupil location and diameter as eye movements occur. Although pupil dilations have a higher signal to noise ratio relative to other measurements like EEG, noise and artefacts can contaminate the signal. These artefacts, such as eye blinks, can be removed in post-hoc analysis using techniques such as independent component analysis (ICA) identification. Other sources of noise can arise during large head movements, so in some experimental contexts head stabilisation techniques such as chin rests, or head tracking algorithms, are employed.

Once pupil responses during repeated task-related trials are imaged, the diameter metric can be processed in a variety of ways. Usually, responses are normalized using z-scores and time aligned to the task event. Artefact removal and filtering to smoothen traces or reduce noise can improve signal fidelity depending on the application. Based on the type of task being conducted, measurements of tonic (windowed) overall average pupil diameter or phasic (time-varying) responses can be extracted. Phasic responses can be useful because the time series contains the peak minimum or maximum pupillary change, the time at which these events occur relative to the task, and consequently the acceleration or rate of change leading to the maximum effect observed.

3) Prior art: Pupillometry has not been used extensively in primary QoE studies, but shows promise as a non-invasive physiological measurement due to pupil diameter correlations with processing load, memory, emotion, and arousal [15]. Early studies demonstrated that simple memory recall of increasing numbers of digits in a sequence corresponds to increasing pupil size [66]. Pupil dilation has also been shown to vary with the strength of memories during retrieval in a paired-association task [67], a further indication of processing load. Retrieval of emotional memories triggers LC activation and corresponding pupillary responses [68], something that may bear relevance to emotional versus neutral states elicited by enhanced QoE. Pupillometry shows that graded LC-NE responses follow evolving strategies for a gambling task with an evolving payoff structure [69]. Effects have also been observed in the context of visual processing. For instance, visual search tasks using distractors that increase difficulty of target recognition lead to increases in pupil diameter [70]. Further, pupillometry has been used to probe linguistic processing, and increased syntactic complexity or effortful listening leads to pupil dilation [71]. Given the importance of intelligibility in QoE quantification, this suggests a meaningful way to assess heightened cognitive load that may accompany difficulty in understanding language in various types of content consumption.

C. Eye Blinking

1) Psychological and physiological processes: Eye blinks occur spontaneously in healthy individuals every few seconds under normal light conditions to clean and lubricate the eye, and reflexively in response to noxious stimuli like contact or extreme light. In addition, frequency of blinking varies with different behaviours. For example, reading, particularly on video terminals, is associated with a decrease in blink rate which may in turn cause dry eyes [72], while active social conversation can lead to increased blink frequency [73]. In fact, blinking behaviour across primate species is correlated with social factors such as group size, indicating a role in vigilance and visual awareness [74]. Further, the physical motions of the eye, and movements of the eyelid, are closely related for spontaneous, voluntary and reflexive behaviours [75]. This suggests that eye blinks and gaze may share common signalling components that may be observed in measurements of the two modalities.

2) Experiment design and analysis: The average length of a blink is 100-400 ms, and is easily detected with a variety of means. Blinking can be imaged at low resolution with

TABLE II: Overview of studies adopting multimodal assessment techniques (PPG: photoplethysmography; RM: respiratory measurement; ST: skin temperature; EM: eye measurement; FT: facial tracking; SR: self-report).

Ref	Modalities									Stimuli & Conditions	Fusion method	Main conclusion
	EEG	ECG	EDA	PPG[2]	RM	ST	EM[1]	FT	SR			
[32]	X	X	X				X		X	Audiovisual sequences (2 video and 2 audio quality levels).	No fusion performed.	Significant correlation of EEG (alpha and theta bands) and blink duration (EM) with quality levels; other measures yield no significant correlation.
[76]	X	X			X				X	H.264 video (2 quality, 3 resolution, 3 audio immersion levels).	Feature fusion. EEG and peripheral features. Classification: 3-class SVM with Gaussian radial basis function kernel.	Classification of low, medium, and high immersiveness level (IL) with 61%, 11% and 94% accuracy, respectively, consistent with self-report.
[77]	X	X								18 clips from different movies to evoke 9 different emotions.	No fusion. ECG features for classification of arousal level, EEG features for valence recognition (positive/negative).	Better performance attained by EEG; moderate correlation between EEG and ECG needs to be explored.
[78]	X		X					X	X	11 movie clips to assess engagement for affective cinema.	Decision fusion. Majority vote over three classifiers. Late-fusion strategy for finding the optimum classifier weights.	All modalities contribute to measuring engagement. Comparable contributions by EEG and EDA (GSR). Best performance with majority voting.
[35]	X		X	X	X	X			X	High and low dynamic range (HDR/LDR) video sequences.	Decision fusion. Classification of EEG and peripheral physiological measurements fused by a weighted product.	Significant differences between LDR and HDR for all physiological measures except RM on a number or QoE variables (quality, interest, naturalness, contrast).
[79]	X			X			X			Stress due to delay in video streaming / navigating online news.	Not clear whether fusion was performed.	Cognitive load (as a proxy for stress level) increases 3% with 2 s delay and 15% with 3 s rebuffering. Distraction measured through GT to explain data anomalies.
[14]	X					X	X		X	Assessment of eye fatigue for 2D/3D video sequences.	No fusion performed, only correlation among measurement modalities and SR is evaluated.	Correlation between blink rate (BR) and SR is highest. Correlation between BR and ST is lowest.
[80]		X	X			X			X	Comparing visual fatigue between a 1 h movie in 2D or 3D.	No fusion performed.	EDA and ST were effective in measuring differences between 2D and 3D, ECG in terms of HR was not.
[47]			X	X		X	X			Audiovisual sequences with asychrony and bit rate variations.	No fusion performed.	Lack of significant effects for all physiological responses, argued to be due to short duration and/or low level of the degradations.

[1]Due to space limitations and the sparsity of methods used, we present EM here as an aggregation of gaze tracking (GT), blink rate (BR), and blink duration (BD).

[2]PPG is typically used to measure heart rate (HR) and/or blood volume pulse (BVP).

conventional eye tracking systems that detect a break in gaze when corneal reflections and image of the pupil are lost, or they can be captured at high-resolution with high-speed videography for greater signal fidelity. Eye blinks are also easily captured via muscular activity using Electromyography (EMG). In EEG measurements, although eye blinks are usually considered a primary source of signal contamination that must be removed, these events are easily recovered from activity in the frontal electrodes around the forehead and temples. Typical attributes that are measured include blink rate (BR) and blink duration (BD).

3) Prior art: Previous studies have demonstrated an association between eye BR and visual discomfort [14]. One study examining QoE viewing on 3D displays showed opposite effects on changes in eye BR and visual discomfort based on whether presented stimuli was in-depth motion, which caused an increase in rate, or planar motion, which resulted in a decreased rate [81]. Other work assessing effects of stereoscopic displays have shown that individuals with slow fusional rates and low fusional limits are more susceptible to visual fatigue [82].

VI. Multimodal Techniques

With the advent of advanced media technologies and related immersive and interactive experiences it is expected that overall QoE cannot be measured with any one physiological modality alone. It is more likely that combining measurements from multiple, complementary physiological modalities will improve inference across a wider range of cognitive processes and thus lead to a deeper insight into the experience of the user. In this section we present some multimodal approaches that aim to capture high-level QoE concepts such as engagement, immersion, and sense of presence. An overview of these approaches is presented in Table II with details on the physiological measurement modalities, stimuli and conditions, fusion methods, and main conclusions. We note that this table is incomplete but used here to illustrate some of the recent findings in the field and the challenges when using multimodal approaches.

The experiences assessed with multimodal approaches are broad and range from traditional applications, such as multimedia quality assessment [32], [47], through more advanced applications, such as assessing visual fatigue for 3D video [14], [80] and tone mapping perception for high dynamic range (HDR) video [35], to higher-level experiences, including immersiveness [76] , emotion [77], stress [79], and engagement [78]. Most studies do not abandon traditional self-reporting but rather include it as a well understood reference. Physiological measurement modalities are often chosen in a redundant and

overlapping fashion to obtain deeper insight into a particular experience and to detect anomalies in measurements. For instance, in [79] unexpected decreases in stress levels could be explained by distraction of participants as indicated through GT measurements. We note that the complementary and redundant information across modalities not only provides deeper insight into QoE but improves the measurement system's robustness to failure.

While simultaneous observations of multiple physiological measurements and correlations thereof provide valuable insight in themselves, true multimodal approaches need to suitably fuse all modalities for quantifying overall QoE. The choice of effective fusion methods is key to successful model design. Two categories of fusion strategies are typically considered: feature fusion (or early integration / early fusion) and decision fusion (or late integration / late fusion). In feature fusion, features are extracted from different modalities and then concatenated into a single feature vector, which is then used as input to a classifier. In decision fusion, feature vectors from each channel are used as inputs to independent classifiers, whose outputs are then combined. Very few works [35], [76], [78] jointly consider and fuse measurements from different modalities, and majority voting appears to be a common fusion strategy. We argue that significantly more investigation into fusion methods, especially biologically inspired ones, is needed to further advance multimodal approaches.

Conclusions drawn from the studies considered in this survey typically report on correlations of physiological measurements with system parameters and self-report. Some widely adopted modalities, such as electroencephalography (EEG) and eye measurement (EM), typically report better agreement than other peripheral measures, such as electrocardiography (ECG) and respiratory measurement (RM) (see for instance findings in [32], [35], [77], [78], [80]). Not all studies are in agreement though, with [32] finding no significant correlation for electrodermal activity (EDA) but [80] finding EDA to be an effective measurement technique. In [47], no significant effects were found for any of the deployed physiological measurements. With so many system, user, and context related factors impacting QoE and yet so little evidence, it is not possible to derive firm arguments for or against any of the physiological methods at this early stage. We can argue though that formalising and standardising physiological measurement methods for QoE may guide experiment design and thus lead to stronger conclusions. The recency of the works presented in this section is evidence of a considerable effort towards a better understanding of multimodal approaches. The modalities listed in Table II show that these efforts go beyond the methods discussed in this survey but include a much broader range of physiological measurements.

VII. Discussion and Future Directions

We provided a broad overview of existing physiological measurement techniques. The number and variety of studies combining traditional psychophysical assessment with physiological measures reveal widespread interest in this emerging research area. The presented results are largely in agreement with each other and with standard psychophysiology literature suggesting that inferences made from physiological measurements are useful in the context of QoE. Below, we discuss some general considerations when using physiology for QoE assessment and provide suggestions for future research.

A. Psychophysiological versus Psychophysical Assessment

As mentioned previously, we do not argue for psychophysiological assessment to replace traditional psychophysical assessment, but to augment it when applicable. Both methodologies have their advantages and disadvantages, but when combined will improve overall insight into QoE. In Table III we summarise our view on the major differences between psychophysiological and psychophysical assessment. We note that with psychophysical assessment we here refer to the widely used self-reporting with quantitative scales, such as Likert ratings.

It is apparent that psychophysiological assessment exhibits several important advantages over traditional psychophysical assessment, including avoidance of subjective rating scales, direct insight into internal physiological processes, potentially instantaneous or better responsiveness, and lowered dependency on the task. On the other hand, psychophysiological assessment typically exhibits increased complexity in experimental design and implementation, signal to noise concerns, and the need for more sophisticated analysis techniques. While these observations may hold in most cases they cannot be generalised to all applications and contexts and are for general guidance only.

B. Advancing QoE Research Through Psychophysiology

We believe that psychophysiology not only contributes to the success of individual experiments, but that it can be instrumental in furthering the field of QoE research. In the existing literature, internal cognitive processes and experiences are widely hypothesised towards definitions of terms such as quality and QoE. Physiological methods can help to improve understanding and validating these internal processes in the media consumer. Toward this end, one strength of physiological measurements that should be exploited is the detection of target stimuli that would stay undetected in self-reporting experiments, for instance, through early pre-conscious ERP components. Capturing such effects can be expected to have a profound impact in the context of extensive media consumption, where a sub-/near-threshold distortion may not be actively perceived but may have a long-term effect on overall QoE including, for instance, visual fatigue. Careful experimental design is needed to identify these thresholds.

Furthermore, studies to date mainly focus on experimental paradigms where participants still perform self-reporting in addition to physiological measurements. Research in this domain should focus on the estimation of conscious responses and measurement of the consumer state. This may lead to more truthful results as the media experience is less disturbed. Towards this end, physiological measurement devices need to be as non-intrusive as possible.

TABLE III: Overview of the differences between conventional psychophysical experiments using a self-reporting scale and psychophysiological experiments in the context of QoE assessment.

	Psychophysical assessment	**Psychophysiological assessment**
Experiment design	Well reported in international standards such as the ones developed by the ITU [83], [84].	Not well understood yet in the context of QoE assessment, and standardisation activities are in early stages.
Implementation	Typically straightforward, involving stimuli presentation and display of a response scale.	Typically more difficult as the stimulus presentation needs to be integrated and accurately synchronised with the physiology measurement device(s).
Scale encoding	The response of an observer is often given on a discrete or continuous scale, such as a Likert scale. It is not well understood how such scales reflect internal perceptual scales of the observers. Methods such as paired comparison can alleviate this problem and are becoming more popular.	No encoding to a scale is needed. Values are understood as a relative measure between conditions rather than absolute numbers.
Response signal quality	The signal quality is strongly influenced by the response scale used (type of scale, granularity, etc.) as well as the participants of the experiment (e.g. experts versus crowdsourcing). Aggregated responses are further impacted by 'between' and 'within' participant differences. These sources of variation can be treated with well established statistical methods (e.g. outlier detection).	The signal is typically noisy, both from the physiological response itself as well as the measurement process (sensor, amplifier, cables, etc.). The latter depends strongly on the quality of the physiological measurement device. Despite the noise, subtle differences in the response can typically be detected over a number of stimulus repetitions.
Nature of responses	Observers provide a conscious response on a given scale. Hence, insight into internal perceptual and cognitive processes is obtained.	Direct insight into internal physiological processes is obtained but no conscious response is obtained.
Task dependency	High-level cognitive processes that lead to conscious response decisions are highly task dependent.	Low-level physiological responses are expected to be less task dependent, but this conjecture needs to be experimentally validated.
Temporal dependencies	Responses are typically retrospective and memory effects may occur for long stimuli.	Responses are largely instantaneous and only dependent on the latency of the physiological response.
Analysis of responses	Statistical analysis techniques, such as the mean opinion score (MOS), significance tests, and outlier detection, are well documented and understood.	Analysis techniques for most physiological measures are complex and often include machine learning and pattern recognition in addition to statistical analysis. These techniques are well documented and understood in domains other than QoE assessment.
Ecological validity	Difficult to achieve as experiments are typically performed in the lab. Even if the study is performed in the observer's usual environment, the concious nature of the response typically breaks the immersion with the task or the content.	In case the psychophysiological measurements can be performed in a non-intrusive way in the observers normal environment, ecological validity could be achieved. Current equipment is largely prohibitive of this.

C. Limitations and Challenges

While we see a clear value in psychophysiology for QoE assessment, there are also a number of limitations and challenges that exist in these methods. Many challenges are inherent to the rapidly evolving nature of advanced media technologies, which increasingly goes beyond passive content into new forms of media centred around virtual and augmented reality and interactive content. Immersive visual environments, such as 360 degree or multiple viewpoint video, can alter assessment targets and create conditions of physical human movement making stable, low-noise recordings difficult to obtain.

The individual physiological difference between humans, which may generate systematic errors between participants or groups thereof, underscores another principle challenge for QoE assessment that is designed to be applicable to a general population. Data processing methods such as spatial filters show promising results in overcoming these systematic inter-participant variances. Screening processes analogous to those used in behavioural experiments may further alleviate this challenge. On the other hand, inter-subject perceptual and physiological differences that can be quantified and tracked could in turn be exploited to tune an enhanced individual experience.

Intrusiveness is considered a strong limitation of physiological measurement techniques. Attaching sensors to subjects can result in the subject feeling a certain degree of discomfort or otherwise change their natural behaviour. Moreover, experiments requiring attached sensors are often considerably more complex, labour intensive and typically require a high level of skill on the part of the experimenter. Such factors can have a direct impact on experiment duration and consequently, on the availability of participants. Nonetheless, it is also important to note that the structure of a psychophysical assessment itself can also modify subject behaviour, whereby a stimulus presentation often deviates substantially from more natural media consumption due to interruptions associated with tasks such as self-reporting or trial structure.

The advent of less intrusive technologies could help mitigate the above concerns and potentially open up new consumer applications in the future. For example, dry-contact EEG sensors and eye trackers that are being integrated into mobile and head-mounted devices will offer the potential to seamlessly track neural states and overt visual attention during an experience, respectively. Non-contact sensing technologies, many of which are beyond the scope of this review, are also likely to make contributions to QoE assessment. Visible light spectra and thermal imaging are actively being explored and may provide correlative hosts of autonomic and cognitive states. For instance, computer vision algorithms for facial expression processing and motion capture are informative for assessing the emotional state and activity of individuals and their interactions.

An important challenge that remains for both psychophysiological and psychophysical QoE assessment is the true measurement of high level cognitive variables. Depending on the paradigm, physiological methods measure stress, attention,

fatigue etc., from which an even higher level cognitive QoE, such as engagement or immersion, needs to be derived. Moreover, different measurement techniques suffer from variable spatial or temporal resolution. Multimodal approaches that combine several physiological measurements into a higher dimensional feature space and which balance sampling resolution will improve overall prediction accuracy. Combining such representations with self-assessment is expected to alleviate this challenge and early work in this direction has shown promising results (see Section VI). Lastly, the robustness of algorithms designed to infer cognitive state will perhaps be most vigorously tested in real-time closed-loop settings. Parallels can perhaps be drawn to recent advances in neural prosthetics and brain-computer interfaces (BCIs) in which an offline-trained algorithm is transitioned to an online dynamical system that is faced with non-stationary statistics, changing and noisy environments, and stability requirements [85] [86].

D. Impact of Physiological Measurement Devices

In psychophysiological research, knowledge is derived about cognitive states from physiological measurements. Compared to traditional QoE assessment methods we are adding an element to the equation that may potentially introduce errors in the conclusions we derive, by essentially quantifying distortions in QoE systems using a device that inherently provides noisy and possibly distorted signals itself. It is therefore important to understand the quality of the physiological response provided by the measurement tool used.

The choice of the device is typically subject to a quality-cost trade-off given the availability of a wide range of consumer, research, and medical grade physiological measurement devices. Numerous wearable sensors and systems have been developed with the main aim to promote the usability of physiological measurements in real-world scenarios. While they usually come at low cost, the capabilities and quality provided are often limited. Research and medical grade physiological measuring devices come with a number of features but are associated with higher costs and often less flexibility. Rigorous studies are needed to fully understand the relationship of a broad range of physiological measurement devices and their impact on QoE assessment.

E. Novel Analysis Techniques

Our literature survey revealed that most studies readily deploy data analysis techniques that are well established in the physiological measurements domain, such as signal magnitude, latency, and frequency analyses. In the context of QoE assessment, these measures have exhibited correlations with psychophysical responses but only moderate accuracy for predicting QoE (see for instance [34], [87]). We argue that this may be due to several reasons, two of which include the fuzziness and the lack of capturing higher order patterns in the physiological responses.

Fuzzy analysis has recently been shown to be of value in image quality assessment [88], [89]. Here, fuzzy regression models predict the psychophysical qualities in fuzzy numbers, which indicate both magnitudes and uncertainties of psychophysical data [90]. As such, they account for the variability of human responses in self-report, both within and across observers. In the context of physiological measures, these techniques may be able to capture the noisiness of the measurement device and, more importantly, the fuzziness of the physiological response, which can strongly vary between humans but also within humans depending on their current state and the given context. Furthermore, large numbers of samples are required for statistical regression to satisfy normal distribution constraints, while fuzzy regression often requires less experimental data [91]. Whether these factors will be of value in physiological assessment remains to be explored.

Machine learning [92] and related techniques such as support vector machines [93] and genetic programming [94] have been successfully used to predict QoE based on psychophysical responses. The power of these techniques is the discovery of patterns in the data that may otherwise be hidden. Recently, deep learning techniques, such as convolutional neural networks, have been shown to discover complex patterns in data at various scales. Not to be overstated, increased computing power and the collection of large numbers of training data has fuelled the recent jump in predictive power of deep nets for computer vision and natural language processing applications. Similarly, as the number of psychophysiological training examples continues to grow, machine learning holds considerable potential to improve physiology-based QoE model prediction and generalization across use cases.

Furthermore, a common technique used in BCIs is to project the data recorded in sensor-space to subspaces by a linear filter that is designed according to a particular criterion. For ERP-based quality assessment, linear discriminant analysis was shown to be a feasible filter, both for feature extraction and for classification, for assessing video quality [19]. In [29] common spatial patterns find a low dimensional subspace to classify the perceived quality of images in an SSVEP experiment. In addition, state-space based modelling has been exploited for trajectory decoding used in neurally-controlled motor BCIs and will likely prove valuable for inferring the time-varying cognitive state in QoE applications [95].

F. Publicly Accessible Psychophysiology Databases

To build sufficient psychophysiological evidence in support of identifying the various factors impacting QoE and to enable effective algorithm development it is essential to have sufficient psychophysiological data. We argue that such evidence needs to be created collectively and made openly accessible to the research community. An overview of some already available databases is provided in Table IV.

True multimodal data sets are provided by Soleymani et al. [96] in their database MAHNOB-HCI and Abadi et al. [97] in their database for decoding user physiological responses to estimate affective multimedia content (DECAF). Besides some of the measures discussed in our survey, several other peripheral measures are included in these databases such as magnetoencephalogram (MEG), electrooculogram (EOG), electromyogram (EMG), respiratory measurements (RM), facial tracking (FT), and skin temperature (ST). Perrin et al. [76],

TABLE IV: Overview of publicly available psychophysiology databases for QoE assessment (N: number of subjects; MEG: magnetoencephalogram; EOG: electrooculogram; EMG: electromyogram; RM: respiration measurement; FT: facial tracking; ST: skin temperature).

Database	N	EEG	ECG	EDA	GT	MEG	EOG	EMG	RM	FT	ST
MAHNOB [96]	27	X	X	X	X				X	X	X
DECAF [97]	30	X	X			X	X	X		X	
Database1 [76]	20	X	X						X		
Database2 [98]	20	X	X						X		
DEAP [99]	32	X								X	
SEED [100]	15	X								X	
Survey of 28 databases [101]	-				X						

[98] published two extensive databases from their study on multimodal QoE assessment of immersive media. Koelstra et al. [99] present DEAP, a multimodal data set for the analysis of human affective states when watching video and Zheng et al. [100] provide the SJTU Emotion EEG Dataset (SEED) from their study on emotion recognition when watching video. Probably the most extensive range of databases exist for GT recordings. Winkler and Subramanian surveyed 28 GT databases for image and video applications and provide an excellent overview of their characteristics [101].

While these data sets are highly valuable for the research community, they provide only small pieces of the overall puzzle. Significantly more data sets need to be made available to fully understand all the factors influencing QoE. In that regard, systematic methodologies are also needed to make compatible and integrate physiological data of different databases to allow for effective analytics and hypothesis testing.

G. Towards Standardisation

Standardisation of physiological methodologies for QoE assessment is critical for researchers to properly design their experiments, allow for cross-laboratory consistency and comparison, and ultimately to further the field of QoE research. There is a lack of such standards but several standardisation activities are currently ongoing.

Several contributions on using physiological measures in QoE have been brought to the International Telecommunication Union (ITU) [102]–[106] with the aim to show the value of using physiological measures to support conventional psychophysical methods. A recommendation is being formed under the working item *P.PHYSIO*. The overarching goals are to enable comparable experimental designs with a focus on the test procedure and the measurement environment. The topic was also presented to the European Telecommunications Standards Institute (ETSI) [107].

IEEE is developing relevant standards with a focus on 3D content. In particular, "P3333.1.1 - Standard for the Quality of Experience (QoE) and Visual Comfort Assessments of Three Dimensional (3D) Contents Based on Psychophysical Studies" was approved as a new standard by the IEEE-SA Standards Board on 26 March 2015, and a working group is currently working on "P3333.1.2 - Standard for the Perceptual Quality Assessment of Three Dimensional (3D) Contents based on Physiological Mechanisms".

The Video Quality Experts Group (VQEG) is an international consortium of partners that work toward a better understanding of video quality perception as well as related psychophysical experiment design and prediction tools [108], [109]. The recently founded Psychophysiological Quality Assessment (PsyPhyQA) project extends VQEG's scope to investigate psychophysiological measurements. The aim of the PsyPhyQA project is to establish novel psychophysiology based techniques and methodologies for video quality assessment and real-time interaction of humans with advanced video communication systems. Specifically, some of the aspects that the project is looking at include the definition of experimental methodologies, development of computational prediction models, and the correlates of psychophysics and psychophysiology. The current focus is on the development of an EEG-based experimental test plan and a cross-lab experiment evaluating professional grade with consumer grade EEG equipment for video quality assessment.

H. Future Impact

We are already seeing a trend of physiological measurements being integrated into modern computing devices, such as heart rate monitors in smart watches and gaze trackers in tablet computers. The wide adoption in mass consumer markets indicates that this trend will like continue for the years to come and even move on to more advanced technologies like smart glasses and other wearables. This may provide an opportunity for continuous QoE monitoring in minimally invasive ways and in the natural environment of the users, thus alleviating concerns about ecological validity of the recorded data. The time is now to lay the ground work in understanding physiological measurements in relation to QoE to be prepared for future deployment in advanced media technologies and services.

VIII. Conclusion

We presented a survey of psychophysiology-based QoE assessment. Previous studies have demonstrated that physiological measurements provide valuable insight into QoE of advanced media technologies. Some measures, such as EEG and gaze tracking, are well established in the existing body of research, while the other measures have been comparatively less thoroughly explored in the context of QoE. Neurophysiological measures as well as eye measurements are typically deployed as primary sources of information whereas measures such as ECG and EDA are often considered to be peripheral measures. In general, there is a consensus that multimodal approaches, integrating a range of physiological and psychological processes, are needed to fully understand QoE of advanced media technologies. We argue that psychophysiology should be used to learn more about the internal cognitive processes to mitigate our assumptions about the quality formation process in humans. The increased body of research as well as recent efforts towards standardisation will lead us closer to this goal.

REFERENCES

[1] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual Evaluation of Speech Quality (PESQ) - A New Method for Speech Quality Assessment of Telephone Networks and Codecs," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 749–752.

[2] S. Winkler, *Digital Video Quality - Vision Models and Metrics*. John Wiley & Sons, 2005.

[3] S. Möller and A. Raake, *Quality of Experience*. Springer, 2014.

[4] G. Ghinea, C. Timmerer, W. Lin, and S. Gulliver, "Mulsemedia: State of the Art, Perspectives, and Challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 11, no. 1, p. 17, 2014.

[5] S. E. Palmer, *Vision science: Photons to phenomenology*. MIT press Cambridge, MA, 1999, vol. 1.

[6] R. A. Andersen, L. H. Snyder, D. C. Bradley, and J. Xing, "Multimodal representation of space in the posterior parietal cortex and its use in planning movements," *Annual review of neuroscience*, vol. 20, no. 1, pp. 303–330, 1997.

[7] D. Senkowski, T. R. Schneider, J. J. Foxe, and A. K. Engel, "Crossmodal binding through neural coherence: implications for multisensory processing," *Trends in neurosciences*, vol. 31, no. 8, pp. 401–409, 2008.

[8] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone, "Modulation of oscillatory neuronal synchronization by selective visual attention," *Science*, vol. 291, no. 5508, pp. 1560–1563, 2001.

[9] M. S. Gazzaniga, *The cognitive neurosciences*. MIT press Cambridge, MA, 2009.

[10] R. Desimone and J. Duncan, "Neural Mechanisms of Selective Visual Attention," *Annual Review of Neuroscience*, vol. 18, pp. 193–222, 1995.

[11] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[12] A. S. Jansen, X. Van Nguyen, V. Karpitskiy, T. C. Mettenleiter, and A. D. Loewy, "Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response," *Science*, vol. 270, no. 5236, p. 644, 1995.

[13] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya, "Visual Attention in Quality Assessment: Theory, Advances, and Challenges," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 50–59, 2011.

[14] J. W. Bang, H. Heo, J. S. Choi, and K. R. Park, "Assessment of eye fatigue caused by 3D displays based on multimodal measurements," *Sensors*, vol. 14, no. 9, pp. 16 467–16 485, 2014.

[15] J. Beatty, "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources," *Psychol Bull*, vol. 91, no. 2, pp. 276–292, 1982.

[16] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[17] H. Berger, "Über das Elektrenkephalogramm des Menschen," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 98, no. 1, pp. 231–254, 1933.

[18] 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," *Journal of Vision*, vol. 15, no. 6, pp. 1–46, 2015.

[19] 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.

[20] S. Bosse, L. Acqualagna, A. 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 Optical Engineering and Applications*. International Society for Optics and Photonics, 2015, p. 959914.

[21] J.-N. Antons, R. Schleicher, S. Arndt, S. Möller, A. Porbadnigk, and G. Curio, "Analyzing Speech Quality Perception Using Electroencephalography," *IEEE J. Select. Topics Signal Proc.*, pp. 721–731, 2012.

[22] J.-N. Antons, A. Porbadnigk, R. Schleicher, B. Blankertz, S. Möller, and G. Curio, "Subjective Listening Tests and Neural Correlates of Speech Degradation in Case of Signal-correlated Noise," in *Proc. of the 129th Convention of the Audio Engineering Society*, 2010, pp. 1–4.

[23] J.-N. Antons, *Neural Correlates of Quality Perception for Complex Speech Signals*. Springer, 2015.

[24] 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," in *Proc. of the fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012, pp. 63–67.

[25] J.-N. Antons, F. Köster, S. Arndt, S. Möller, and R. Schleicher, "Changes of Vigilance Caused by Varying Bit Rate Conditions," in *Proc. of the fifth International Workshop on Quality of Multimedia Experience*. IEEE, 2013, pp. 148–151.

[26] S. Arndt, J. Antons, R. Gupta, K. Laghari, R. Schleicher, S. Möller, and T. Falk, "Subjective Quality Ratings and Physiological Correlates of Synthesized Speech," in *Proc. of the IEEE Fifth International Workshop on Quality of Multimedia Experience*. IEEE, 2013, pp. 152–157.

[27] K. Laghari, R. Gupta, S. Arndt, J. Antons, S. Möller, and T. Falk, "Characterization of Human Emotions and Preferences for Text-to-Speech Systems Using Multimodal Neuroimaging Methods," in *Proc. of the IEEE 27th Canadian Conference on Electrical and Computer Engineering*, 2014, pp. 1–5.

[28] 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 *Proc. of the IEEE Int. Conf. Image Process.* IEEE, 2014, pp. 1987–1991.

[29] 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)," *Journal of Neural Engineering*, vol. 12, no. 2, p. 026012, 2015.

[30] S. Arndt, *Neural Correlates of Quality During Perception of Audiovisual Stimuli*. Springer, 2015.

[31] S. Arndt, J.-N. Antons, R. Schleicher, S. Möller, and G. Curio, "Using Electroencephalography to Measure Perceived Video Quality," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 3, pp. 366–376, 2014.

[32] S. Arndt, J.-N. Antons, R. Schleicher, and S. Möller, "Using Electroencephalography to Analyze Sleepiness Due to Low-Quality Audiovisual Stimuli," *Signal Processing: Image Communication*, vol. 42, pp. 120–129, 2016.

[33] J. Donley, C. Ritz, and M. Shujau, "Analysing the Quality of Experience of Multisensory Media from the Measurements of Physiological Responses," in *Proc. of the IEEE International Conference on Wireless Communications and Signal Processing*, 2015, pp. 1–6.

[34] S. Arndt, J. Radun, J. Antons, and S. Möller, "Using Eye Tracking and Correlates of Brain Activity to Predict Quality Scores," in *Proc. of the IEEE Sixth International Workshop on Quality of Multimedia Experience*, 2014, pp. 281–285.

[35] S.-E. Moon and J.-S. Lee, "Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 236–247, 2015.

[36] J. Beyer, R. Varbelow, J.-N. Antons, and S. Möller, "Using Electroencephalography and Subjective Self-Assessment to Measure the Influence of Quality Variations in Cloud Gaming," in *Proc. of the IEEE Seventh International Workshop on Quality of Multimedia Experience*, 2015, pp. 1–6.

[37] E. Kroupi, P. Hanhart, J. Lee, M. Rerabek, and T. Ebrahimi, "EEG Correlates During Video Quality Perception," in *Proc. of the 22nd European Signal Processing Conference*. IEEE, 2014, pp. 2135–2139.

[38] 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.

[39] 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.

[40] S. Arndt, K. Brunnström, E. Cheng, U. Engelke, S. Möller, and J.-N. Antons, "Review on using physiology in quality of experience," *Electronic Imaging*, vol. 2016, no. 16, 2016.

[41] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek, "Brain-Computer Interfacing for Multimedia Quality Assessment," in *Proc. of the 2016 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, 2016.

[42] R. Gupta, K. Laghari, S. Arndt, R. Schleicher, S. Möller, and D. O'Shaughnessy, "Using fNIRS to Characterize Human Perception of TTS System Quality, Comprehension, and Fluency: Preliminary Findings," in *Proc. of the International Workshop on Perceptual Quality of Systems*, 2013, pp. 1–6.

[43] Y. Du, D. Gutu, K. Shibata, Y. Inazumi, and Y. Horita, "A Study of the Relationship Between Image Quality and Cerebral Blood Flow," in *Proc. of the Biomedical Circuits and Systems Conference*. IEEE, 2014, pp. 21–24.

[44] E. Kroupi, P. Hanhart, J.-S. Lee, M. Rerabek, and T. Ebrahimi, "Predicting subjective sensation of reality during multimedia consumption based on EEG and peripheral physiological signals," in *Proc. of the International Conference in Multimedia and Expo*, 2014, pp. 1–6.

[45] J. Pan and W. J. Tompkins, "A Real-time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 230–236, 1985.

[46] M. Barreda-Angeles, R. Pepion, E. Bosc, P. Le Callet, and A. Pereda-Banos, "Exploring the Effects of 3D Visual Discomfort on Viewers' Emotions," in *Proc. of the IEEE International Conference on Image Processing*, 2014, pp. 753–757.

[47] J. Lassalle, L. Gros, and G. Coppin, "Combination of Physiological and Subjective Measures to Assess Quality of Experience for Audiovisual Technologies," in *Proc. of Int. Workshop on Quality of Multimedia Experience*. IEEE, 2011, pp. 13–18.

[48] G. M. Wilson and M. A. Sasse, "Do Users Always Know What's Good for Them? Utilising Physiological Responses to Assess Media Quality," in *People and Computers XIV Usability or Else!* Springer, 2000, pp. 327–339.

[49] U. Engelke, A. Maeder, and H. Zepernick, "Visual Attention Modelling for Subjective Image Quality Databases," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, 2009, pp. 1–6.

[50] P. Le Callet and E. Niebur, "Visual Attention and Applications in Multimedia Technologies," *Proc. of the IEEE Inst Electr Electron Eng*, vol. 101, no. 9, pp. 2058–2067, 2013.

[51] U. Engelke, R. Pepion, P. Le Callet, and H. Zepernick, "Linking Distortion Perception and Visual Saliency in H.264/AVC Coded Video Containing Packet Loss," in *Proc. of the SPIE/IEEE Int. Conf. on Visual Communications and Image Processing*. International Society for Optics and Photonics, 2010, p. 774406.

[52] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Do Video Coding Impairments Disturb the Visual Attention Deployment?" *Signal Processing: Image Communication*, vol. 25, no. 8, pp. 597–609, 2010.

[53] H. Liu, U. Engelke, J. Wang, P. Le Callet, and I. Heynderickx, "How Does Image Content Affect the Added Value of Visual Attention in Objective Image Quality Assessment?" *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 355–385, 2013.

[54] J. Lee, F. de Simone, and T. Ebrahimi, "Influence of Audio-Visual Attention on Perceived Quality of Standard Definition Multimedia Content," in *Proc. of Int. Workshop on Quality of Multimedia Experience*, 2009, pp. 13–18.

[55] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you Gaze on an Image Affect your Perception of Quality? Applying Visual Attention to Image Quality Metric," in *Proc. of the IEEE Int. Conf. on Image Processing*, vol. 2, 2007, pp. 169–172.

[56] U. Engelke, M. Barkowsky, P. Le Callet, and H. Zepernick, "Modelling Saliency Awareness for Objective Video Quality Assessment," in *Proc. of the Int. Workshop on Quality of Multimedia Experience*, 2010, pp. 212–217.

[57] M. Barkowsky, J. Bialkowski, B. Eskofier, R. Bitto, and A. Kaup, "Temporal Trajectory Aware Video Quality Measure," *IEEE J. Select. Topics Signal Proc.*, vol. 3, no. 2, pp. 266–279, 2009.

[58] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt Visual Attention for Free-viewing and Quality Assessment Tasks Impact of the Regions of Interest on a Video Quality Metric," *Signal Processing: Image Communication*, vol. 25, pp. 547–558, 2010.

[59] H. Liu and I. Heynderickx, "Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.

[60] U. Engelke, H. Liu, J. Wang, P. Le Callet, I. Heynderickx, H.-J. Zepernick, and A. Maeder, "Comparative Study of Fixation Density Maps," *IEEE Trans. on Image Processing*, vol. 22, no. 3, pp. 1121–1133, 2013.

[61] G. Aston-Jones and J. Cohen, "An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance," *Annu. Rev. Neurosci.*, vol. 28, pp. 403–450, 2005.

[62] C. Varazzani, A. San-Galli, S. Gilardeau, and S. Bouret, "Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-Off: a Direct Electrophysiological Comparison in Behaving Monkeys," *J. Neurosci*, vol. 35, no. 20, pp. 7866–7877, 2015.

[63] S. Joshi, Y. Li, R. Kalwani, and J. I. Gold, "Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex," *Neuron*, vol. 89, no. 1, pp. 221–234, 2016.

[64] H. S. Thompson, R. H. Kardon, and I. E. Loewenfeld, "Irene E. Loewenfeld, PhD Physiologist of the pupil," *J Neuroophthalmol*, vol. 26, no. 2, pp. 139–148, 2006.

[65] T. Yoshitomi, Y. Ito, and H. Inomata, "Adrenergic Excitatory and Cholinergic Inhibitory Innervations in the Human Iris Dilator," *Experimental Eye Research*, vol. 40, no. 3, pp. 453–459, 1985.

[66] D. Kahneman and J. Beatty, "Pupil Diameter and Load on Memory," *Science*, vol. 154, no. 3756, pp. 1583–1585, 1966.

[67] H. van Rijn, J. Dalenberg, J. Borst, and S. A. Sprenger, "Pupil Dilation Co-Varies with Memory Strength of Individual Traces in a Delayed Response Paired-Associate Task," *PLoS ONE*, vol. 7, no. 12, p. e51134, 2012.

[68] V. Sterpenich, A. D'Argembeau, M. Desseilles, E. Balteau, G. Albouy, G. Vandewalle, C. Degueldre, A. Luxen, F. Collette, and P. Maquet, "The Locus Ceruleus is Involved in the Successful Retrieval of Emotional Memories in Humans," *J. Neurosci*, vol. 26, no. 28, pp. 7416–7423, 2006.

[69] M. Jepma and S. Nieuwenhuis, "Pupil Diameter Predicts Changes in the Exploration-exploitation Trade-off Evidence for the Adaptive Gain Theory," *Journal of Cognitive Neuroscience*, vol. 23, no. 7, pp. 1587–1596, 2011.

[70] G. Porter, T. Troscianko, and I. D. Gilchrist, "Effort during Visual Search and Counting: Insights from Pupillometry," *The Quarterly Journal of Experimental Psychology*, vol. 60, no. 2, pp. 211–229, 2007.

[71] A. A. Zekveld, S. E. Kramer, and J. M. Festen, "Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility," *Ear Hear*, vol. 31, no. 74, pp. 480–490, 2010.

[72] K. Tsubota and K. Nakamori, "Dry Eyes and Video Display Termi-

nals," *New England Journal of Medicine*, vol. 328, no. 8, pp. 584–584, 1993.

[73] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of Blink Rate in Normal Subjects," *Movement Disorders*, vol. 12, no. 6, pp. 1028–1034, 1997.

[74] H. Tada, Y. Omori, K. Hirokawa, H. Ohira, and M. Tomonaga, "Eyeblink Behaviors in 71 Species of Primates," *PLoS ONE*, vol. 8, no. 5, p. e66018, 2013.

[75] L. J. Bour, M. Aramideh, and B. W. de Visser, "Neurophysiological Aspects of Eye and Eyelid Movements During Blinking in Humans," *J. Neurophysiol.*, vol. 83, no. 1, pp. 166–176, 2000.

[76] A.-F. N. M. Perrin, H. Xu, E. Kroupi, M. Řeřábek, and T. Ebrahimi, "Multimodal Dataset for Assessment of Quality of Experience in Immersive Multimedia," in *Proc. of the 23rd Annual ACM Multimedia Conference*. ACM, 2015, pp. 1007–1010.

[77] N. Ramzan, S. Palke, T. Cuntz, R. Gibson, and A. Amira, "Emotion recognition by physiological signals," *Electronic Imaging*, vol. 2016, no. 16, pp. 1–6, 2016.

[78] M. Abadi, J. Staiano, A. Cappelletti, M. Zancanaro, and N. Sebe, "Multimodal Engagement Classification for Affective Cinema," in *Proc. of the Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 411–416.

[79] Ericsson, "Ericsson Mobility Report, Mobile World Congress Edition," 2016.

[80] C. Kim, S. Park, M. Won, M. Whang, and E. Lee, "Autonomic Nervous System Responses Can Reveal Visual Fatigue Induced by 3D Displays," *Sensors*, vol. 13, no. 10, pp. 13 054–13 062, 2013.

[81] J. Li, M. Barkowsky, and P. Le Callet, "Visual Discomfort is not Always Proportional to Eye Blinking Rate: Exploring Some Effects of Planar and In-Depth Motion on 3DTV QoE," in *Proc. of Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2013, pp. 1–6.

[82] D. Kim, S. Choi, S. Park, and K. Sohn, "Stereoscopic Visual Fatigue Measurement Based on Fusional Response Curve and Eye-Blinks," in *Proc. of the 17th International Conference on Digital Signal Processing*. IEEE, 2011, pp. 1–6.

[83] ITU-R Recommendation BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures," *International Telecommunication Union, Geneva*, 2002.

[84] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," *International Telecommunication Union, Geneva*, 1999.

[85] R. A. Andersen, E. J. Hwang, and G. H. Mulliken, "Cognitive neural prosthetics," *Annual review of psychology*, vol. 61, p. 169, 2010.

[86] V. Gilja, C. A. Chestek, I. Diester, J. M. Henderson, K. Deisseroth, and K. V. Shenoy, "Challenges and opportunities for next-generation intracortically based neural prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 7, pp. 1891–1899, 2011.

[87] A. Verdejo, K. Moor, I. Ketyko, K. Nielsen, J. Vanattenhoven, T. Pessemier, W. Joseph, L. Martens, and L. Marez, "QoE estimation of a Location-Based Mobile Game Using on-Body Sensors and QoS-Related Data," in *Proc. of the Wireless Days Conference*, 2010, pp. 20–22.

[88] K. Chan and U. Engelke, "Fuzzy Regression for Perceptual Image Quality Assessment," *Engineering Applications of Artificial Intelligence*, vol. 43, pp. 102–110, 2015.

[89] ——, "Varying Spread Fuzzy Regression for Affective Quality Estimation," *IEEE Transactions on Fuzzy Systems*, 2016.

[90] H. Tanaka, S. Uejima, and K. Asai, "Linear Regression Analysis with Fuzzy Model," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 12, pp. 903–907, 1982.

[91] K. Kim, H. Moskowitz, and M. Koksalan, "Fuzzy versus Statistical Linear Regression," *European Journal of Operational Research*, vol. 92, pp. 417–434, 1996.

[92] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a Predictive Model of Quality of Experience for Internet Video," in *Proc. of the annual conference of the ACM Special Interest Group on Data Communication*, 2013, pp. 339–350.

[93] R. A. Farrugia and C. J. Debono, "A Support Vector Machine Approach for Detection and Localization of Transmission Errors Within Standard H.263++ Decoders," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1323–1330, 2009.

[94] N. Staelens, D. Deschrijver, E. Vladislavleva, B. Vermeulen, T. Dhaene, and P. Demeester, "Constructing a No-reference H.264/AVC Bitstream Based Video Quality Metic Using Genetic Programming Based Symbolic Regression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 8, pp. 1322–1333, 2013.

[95] G. H. Mulliken, S. Musallam, and R. A. Andersen, "Decoding trajectories from posterior parietal cortex ensembles," *The Journal of Neuroscience*, vol. 28, no. 48, pp. 12 913–12 926, 2008.

[96] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[97] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.

[98] A.-F. N. M. Perrin, M. Rerabek, and T. Ebrahimi, "Towards prediction of sense of presence in immersive audiovisual communications," in *Human Vision and Electronic Imaging (HVEI) 2016*.

[99] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[100] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[101] S. Winkler and R. Subramanian, "Overview of eye tracking datasets," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*. IEEE, 2013, pp. 212–217.

[102] ITU-T Contribution COM 12-039, "Investigating the Subjective Judgement Process Using Physiological Data," *International Telecommunication Union, Geneva*, 2013.

[103] ITU-T Contribution COM 12-112, "Using Physiological Data for Assessing Variations of the Cognitive State Evoked by Quality Profiles," *International Telecommunication Union, Geneva*, 2013.

[104] ITU-T Contribution COM 12-103, "Using Physiological Data for Assessing Subjective Video Quality Ratings," *International Telecommunication Union, Geneva*, 2013.

[105] ITU-T Contribution COM 12-202, "Using Physiological Data for Assessing the Audiovisual Quality of Longer Stimuli," *International Telecommunication Union, Geneva*, 2014.

[106] ITU-T Contribution COM 12-307, "Structure Proposal for the Work Item on the Use of Physiological Measures as an Additional Test Method for Speech Quality Assessment (P.PHYSIO)," *International Telecommunication Union, Geneva*, 2016.

[107] European Telecommunications Standards Institute, "Using Electroencephalography to Gain Insights into the Perception of Different Levels of Quality," *European Telecommunications Standards Institute, Vienna*, 2015.

[108] K. Brunnström, D. Hands, F. Speranza, and A. Webster, "VQEG Validation and ITU Standardization of Objective Perceptual Video Quality Metrics [Standards in a Nutshell]," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 96–101, 2009.

[109] Q. Huynh-Thu, A. Webster, K. Brunnström, and M. Pinson, "VQEG: Shaping Standards on Video Quality," in *Proc. of the 1st International Conference on Advanced Imaging*, 2015, pp. 1–4.

Ulrich Engelke (S'05-M'10-SM'13) received a PhD degree in Telecommunications from the Blekinge Institute of Technology, Sweden, in 2010. He is currently a Senior Research Scientist in Cognitive Informatics at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Hobart, Australia. Ulrich is co-chair of the Psychophysiology-based Quality Assessment (PsyPhyQA) project of the Video Quality Experts Group (VQEG). His research interests are within quality of experience, visual analytics, and perceptual imaging.

Daniel P Darcy received a PhD in neuroscience from the University of California, San Diego and was a postdoctoral fellow at the University of California, San Francisco. His research interests include synaptic physiology, neurogenesis, and plasticity. He is currently Staff Scientist at Dolby Labs in San Francisco, CA.

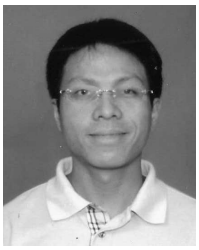
Grant H Mulliken received the Ph.D. degree in Computation and Neural Systems from the California Institute of Technology, Pasadena. Currently, he is a Senior Staff Scientist at Dolby Laboratories, San Francisco. Previously, he was a research scientist in Neuroscience at the Massachusetts Institute of Technology, Boston, MA. His research interests cover neural coding and decoding, sensorimotor control and visual attention, brain-machine interfaces, neurally inspired VLSI and signal processing, as well as machine learning and intelligent systems.

Sebastian Bosse is with the Machine Learning group and the Image & Video Coding group at Fraunhofer Institute for Telecommunications - Heinrich Hertz Institute, Berlin, Germany. He 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 current research interests include video compression, human visual perception and its neural correlates, and machine learning

Maria Martini (S'99-M'02-SM'07) is a Professor in the Faculty of Science, Engineering and Computing in Kingston University, London. She received the Laurea in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002. Her research interests include quality of experience assessment and provision in wireless multimedia networks, 2D/3D error resilient video, 2D/3D video quality assessment, and medical applications.

Sebastian Arndt studied Computer Science at Technische Universität Berlin (TUB), and University of Oklahoma (USA), and received his diploma in 2010. He received his doctoral degree (Dr.-Ing.) in 2015 from TUB with the thesis title Neural Correlates of Quality During Perception of Audiovisual Stimuli. Since 2016 he is a researcher at NTNU (Norwegian University of Technology and Science). His current research focuses on integrating social media and real-time graphics in the broadcasting industry.

Jan-Niklas Antons is a Senior Research Scientist at the Quality and Usability Lab of the Technische Universität Berlin, Germany. He received his diploma in psychology in 2008 from the Technische Universität Darmstadt and a Doctor-of-Engineering degree in 2014 from the Technische Universität Berlin. He was visiting researcher at INRS-EMT, Canada (2012) and NTNU, Norway (2014). His research interests are in Quality-of-Experience evaluation and its physiological correlates with an emphasis on media transmissions and human-machine-interaction.



Kit Yan Chan is a Senior Lecturer in the Department of Electrical and Computer Engineering, Curtin University, Australia. He received his PhD in Computing in London South Bank University, U.K., in 2006. He was a full time researcher in The Hong Kong Polytechnic University (2004-2009) and Curtin University (2009-2013). His research interests include machine learning and its applications to perceptual quality evaluation, new product development, manufacturing process design.

Naeem Ramzan (S'04-M'08-SM'13) received the M.Sc. in telecommunication from University of Brest, France and PhD in electronic engineering from Queen Mary University of London in 2004 and 2008 respectively. He is currently a Reader in University of the West of Scotland. He has investigated/co-investigated/participated several projects funded by EU and UK research councils. He has authored or co-authored over 90 research publications He is senior member of IEEE, and fellow of HEA.

Kjell Brunnström , Ph.D., is a Senior Scientist at Acreo Swedish ICT AB and Adjunct Professor at Mid Sweden University. He is Co-chair of the Video Quality Experts Group (VQEG). He has written about 100 peer-reviewed scientific journal articles and international conference papers as well as served as reviewer for scientific journals and international conferences. His research interests are in Quality of Experience for visual media in particular video and display quality assessment.