EEG CORRELATES DURING VIDEO QUALITY PERCEPTION.

Eleni Kroupi[1], *Philippe Hanhart*[1], *Jong-Seok Lee*[2], *Martin Rerabek*[1], *and Touradj Ebrahimi*[1]

[1]Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland

[2]School of Integrated Technology
Yonsei University
162-1 Songdo-dong, Yeonsu-gu, 406-840 Incheon, Republic of Korea

ABSTRACT

Understanding Quality of Experience (QoE) in various multimedia contents is still challenging. In this paper, we investigate the way QoE affects brain oscillations captured by electroencephalography (EEG). In particular, sixteen subjects watched 2D and 3D videos of various quality levels while their EEG signals were recorded, and were asked to provide their self-assessed perceived quality ratings for each video. EEG signals were decomposed into six frequency bands, namely theta, alpha, beta low, beta middle, beta high and gamma bands. The results revealed frontal asymmetry patterns in the alpha band, which correspond to right frontal activation when perceived quality is low. This finding implies that perceived high quality may be related to positive emotional processes.

***Index Terms*—** EEG, QoE, frontal asymmetry

1. INTRODUCTION

User Quality of experience (QoE) has been considered as an important issue for various image and video applications. It is necessary to understand how humans perceive quality from visual stimuli, which can be potentially exploited for developing and optimizing image and video processing algorithms.

Subjective quality assessment is usually employed to measure users' perceived quality. Typically, a number of subjects are asked to rate given stimuli on a pre-defined rating scale. The gathered quality scores are processed to obtain mean opinion scores (MOS) or differential MOS (DMOS) for the stimuli. While this requires subjects' explicit responses, attempts towards implicit monitoring of human quality perception in the brain have been made recently. For instance, 3D visual fatigue was detected through electroencephalogram (EEG) response [1] or through cortical activities measured by fMRI [2]. Also, an approach of direct estimation of 2D video quality perception through EEG evoked-related potentials (ERPs) has been investigated [3]. In general, such direct monitoring approaches require no explicit response of subjects and can, thus, exclude subjective biases that may occur during the rating procedure. However, implicit QoE approaches are still in their infancy and further research needs to be performed to better understand the nature of the recorded neural signals in the viewpoint of QoE. Questions such as what can or cannot be inferred from the EEG signals in terms of QoE, and which QoE-related features can be extracted from the EEG signals have not been fully investigated. For instance, although EEG frequency bands, such as theta, alpha, beta, and gamma have been extensively used in various emotion and cognitive-related applications, their importance for QoE has not been adequately investigated.

In this paper, we investigate how QoE of 2D and 3D videos affects brain oscillations through EEG analysis. The aim is to understand which frequency components in the EEG signals that emanate from various brain regions are related to video quality perception. We present our experiments in which 2D and 3D videos with various quality levels are shown to subjects while their EEG signals are recorded. At the end of each video stimulus, subjects provide their self-assessed ratings in terms of QoE. The neural activity patterns in various EEG frequency bands and brain regions are investigated in relation to perceived quality.

The rest of the paper is organized as follows. In the next section, we describe how our experiments were conducted. Section 3 presents the analyses of the subjective ratings and EEG signals. Finally, conclusions are given in Section 4.

The research leading to these results has been performed in the framework of Swiss National Foundation for Scientific Research (FN 200020-132673-1, FN 200020-149259-1, FN IZK0Z2-150904, and FN 200020-143696-1) and the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning, Korea (2013R1A1A1007822). This work was also partially supported by the COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services - QUALINET

Table 1. Characteristics of the contents used in our experiments.

Content	Description and characteristics
Training	Rock band playing at the Auditorium Stravinski. Dark. Bright spots. Shot from the back of the auditorium.
Jazz	Jazz band playing at the Funky Claude's Lounge at the Opening Party. Wide shot.
Rock	Rock band playing at the Auditorium Stravinski. Dark. Bright spots. Shot from the back of the auditorium.
Stage	MJF general manager on stage introducing the next artist. Very dark. In French. Wide shot.
Speech1	MJF general manager giving a speech at the Opening Party. In French. Mid shot.
Speech2	Speech at the Opening Party. In French. Mid shot.
Outdoor	Crowd walking on the street near the lake. Lot of depth. Wide shot.
Interview	Interview of Quincy Jones. Medium close up.

2. DATA COLLECTION

2.1. Participants

Sixteen subjects (5 females, 11 males, between 19 and 30 years old with an average of 23.8 yrs old) participated in our experiments. All subjects were screened for correct visual acuity (no errors on 20/30 line), color vision and stereo vision using Snellen, Ishiara and Randot charts, respectively. All participants provided written consent forms. Before each experiment, a training session was organized to allow participants to familiarize with the procedure and the assessment sheets. The content shown in the training session was selected by the experimenters in order to include various examples.

2.2. Audio-visual stimuli

The dataset comprised eight video contents: one for the training and seven for the tests. All contents were shot during the 2012 edition of the Montreux Jazz Festival (MJF) (protected by copyright), with two RED SCARLET-X cameras mounted on a Genus Hurricane Rig. All video sequences were recorded in REDCODE RAW (R3D) format, DCI 4K resolution (4096×2160 pixels), at 25 fps, and had a duration of about one minute. Stereo audio was recorded in PCM format, sampled at 48 kHz, 24 bits. Table 1 describes the contents and their characteristics. The recorded video sequences were cropped and downsampled to Full HD resolution (1920×1080 pixels) and then compressed with H.264/ MPEG-4 AVC. Two different quantization parameters (QP) were selected: QP=2 for high quality (HQ) and QP=35 for low quality (LQ). For each content, four different versions were considered: 2D HQ, 3D HQ, 2D LQ, and 3D LQ, leading to a total of 28 video sequences, 14 of which in 2D and 14 in 3D.

2.3. Monitor, sound system and environment

To display the video stimuli, a HD 46" Hyundai S465D polarized stereoscopic monitor was used. The laboratory setup was controlled in order to ensure the reproducibility of results by avoiding involuntary influence of external factors. The test room was equipped with a controlled lighting system with a 6500K color temperature and an ambient luminance at 15% of the maximum screen luminance. For the audio playback, the PSI A14-M professional studio full range speakers were used.

2.4. Physiological signal acquisition

The EEG signals were recorded at 250 Hz from 256 electrodes placed at the standard positions on the scalp. An EGI's Geodesic EEG System (GES) 300 was used to record, amplify, and digitalize the EEG signals while the participants were watching the stimuli.

2.5. Experimental protocol

The participants were seated at a distance roughly 1.8 meters from the stereoscopic monitor, as suggested by [4]. All video sequences were viewed with 3D glasses. The experiments were conducted in three sessions. Nine video sequences, randomly chosen, were presented in the first and second sessions, and ten in the last one, leading to a total of 28 video sequences, and thus, to a total of 28 trials.

Each trial consisted of a ten-second baseline period and an one-minute stimulus period. The EEG signals recorded during the baseline period were used to remove stimulus-unrelated variations from the signals obtained during the stimulus period. During the baseline period, the subjects were instructed to remain calm and focus on a 2D white cross on a black background presented on the screen in front of them. Once the baseline period was over, a video sequence was randomly selected and presented. After the video sequence was over, the subjects were asked to provide their self-assessed ratings for the particular video sequence without any restriction in time, following the Absolute Category Rating (ACR) evaluation methodology [5].

Regarding the self-assessed ratings, subjects were asked to evaluate their experience from each video sequence in terms of perceived overall quality. A 9-point rating scale was used that ranged from 1 to 9, with 1 representing "low" perceived overall quality and 9 "high" perceived overall quality.

Once a trial was over, the next trial started by recording the signals during the next baseline period, and the next video sequence was randomly selected and presented. The procedure was repeated until all 28 video sequences had been presented and rated. Although each experiment lasted for almost two hours, including the training and set up, the subjects did not report fatigue.

3. ANALYSIS

3.1. Subjective rating analysis

First, a normality test was performed using the Pearson's chi-squared test to determine if the subjective ratings are well-modeled by a normal distribution. Results showed that, for each individual condition, the ratings of the different subjects were normally distributed. Then, the mean opinion score (MOS) and associated 95% confidence interval (CI) were computed for each test stimulus, to represent explicit estimates of perceived overall quality. Figure 1 shows the resulting MOS and CI. As it can be observed, high quality sequences resulted in significantly higher ratings when compared to their corresponding low quality versions, since the CIs for HQ and LQ stimuli do not overlap, expect for content *Outdoor*. Similarly, no significant difference can be observed between 3D and 2D stimuli, which means that the rendering mode had no significant influence. However, even though the same QPs were used for all sequences, the MOSs associated with high and low quality stimuli vary across the different contents, as contents have different spatial and temporal characteristics.

To investigate quantitatively whether the objective factors, such as the rendering mode, actual quality level, and content have a significant influence on perceived quality, an ANOVA analysis was performed on the subjective ratings. In particular, the null hypothesis was that the rendering mode, actual quality level, and content do not influence perceived overall quality. Table 2 reports the ANOVA results. Results show that the null hypothesis was rejected for content and actual quality level, $p < 0.001$, indicating that these individual factors had a significant impact on perceived overall quality. However, the rendering mode had no significant impact, $p = 0.069$. The interaction between content and actual quality level, $p = 0.049$, as well as between rendering mode and quality level, $p = 0.020$, were found to have a significant impact on perceived overall quality. However, the other interactions were not significant, $p > 0.3$.

3.2. EEG signal analysis

In this section, the pre-processing steps to remove the artifacts, the feature extraction methods and the EEG correlates are presented.

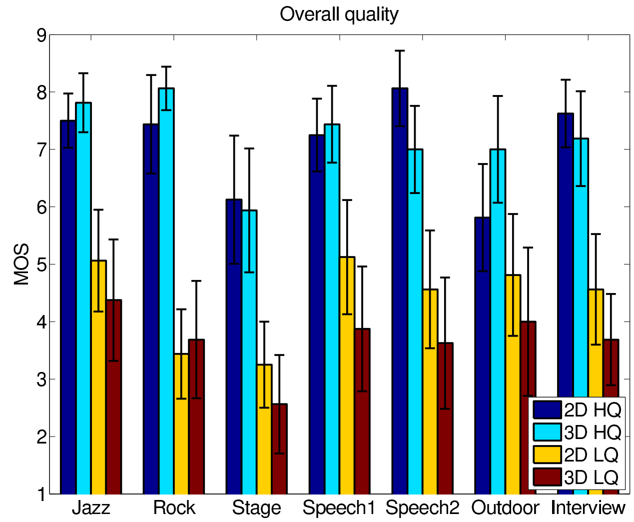


Fig. 1. Mean opinion scores.

Table 2. ANOVA analysis.

Source	df	F	p
content	6	5.968	< 0.001
rendering	1	3.323	0.069
quality	1	330.4	< 0.001
content*rendering	6	1.201	0.305
content*quality	6	2.131	0.049
rendering*quality	1	5.493	0.020
content*rendering*quality	6	0.637	0.701

3.2.1. Preprocessing

EEG electrodes in which muscle activity was discernible were rejected manually, leading to a total 216 electrodes for processing and analysis. EEG signals were filtered between 3-47 Hz using a third-order Butterworth filter, in order to remove electrooculogram (EOG) and electromyogram (EMG) artifacts. Remaining artifacts were removed by cubic interpolation. EEG signals were initially referenced to the Cz electrode and re-referenced to the common average. EEG electrodes were manually clustered into ten clusters depending on which brain region they belong to, namely left and right frontal, left and right central, left and right parietal, left and right temporal, and left and right occipital ones.

Only the last 45 seconds of all signals were used in our analysis, considering that stabilization and adaptation to 3D contents may take some time.

3.2.2. Feature extraction

Regarding the EEG signals, the power spectral density (PSD) of all signals was extracted for frequencies between 4 and 47 Hz, using the Welch's method with windows of 128 samples. The mean trial power was then divided by the mean baseline power, in order to extract the power changes without considering the stimulus-unrelated variations. These power changes were captured for different frequency bands, namely theta (4-7 Hz), alpha (8-12 Hz), beta low (13-16 Hz), beta middle (17-20 Hz), beta high (21-29 Hz), and gamma (30-47 Hz) bands.

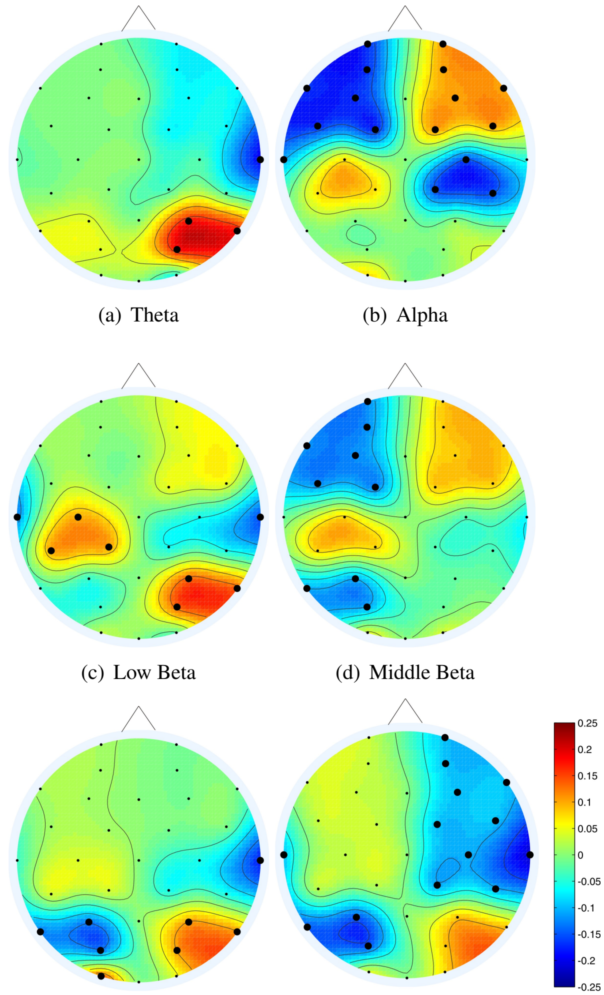
3.2.3. *EEG correlates*

In order to estimate the EEG correlates, the medians of all PSD values that emanate from the same brain regions were extracted to represent the central tendency in terms of power for each brain region and frequency band, across subjects. Then the Spearman correlations were estimated between the median power values and the subjective ratings, for each frequency band. A t-test was applied to the correlation values to explore their level of significance with respect to zero correlation. The results are presented in Figure 2. For visualization purposes, each of the ten brain regions contains several dots that correspond to the electrodes that belong to a specific brain region based on the 10-20 International System with 32 electrodes (32 dots in total). The largest black dots declare the significant correlations for each frequency band and brain region.

Regarding the colors of this plot, the red color corresponds to positive correlation whereas the blue color corresponds to negative correlation between power and subjective ratings. It has been established that low alpha power indicates brain activation, whereas high alpha power brain deactivation [6]. Thus, it seems that the left frontal brain region is activated when overall perceived quality is high, and the right frontal one when overall perceived quality is low. Such frontal asymmetry has been shown to convey emotional information. Specifically, right frontal alpha power activation is related to withdrawal affective processes, whereas left frontal activation to approach-oriented ones [7]. Thus, our outcome corroborates with neurophysiology as overall quality perception contains underlying affective information, which follows the same pattern. This finding implies that indeed subjects in this study rated their perceived quality by taking into account how pleasant or annoying quality was for them.

Moreover, high gamma power corresponds to high brain activity, thus the fact that the right frontal and central cortices are highly activated when perceived quality is low, also corroborates with the above neurophysiological studies. This indicates again that low perception of quality is related to negative emotions.

It has been also established that beta band is highly associated with cognition, and reflects emotional characteristics [8]. In accordance with [8], we also found significantly higher beta on the right parietal lobe compared to the left one for positive emotional tasks (high perceived quality in our case). Hence, cognitive and emotional processes seem to take place during quality perception from 2D and 3D videos.



(a) Theta (b) Alpha (c) Low Beta (d) Middle Beta (e) High Beta (f) Gamma

Fig. 2. EEG correlates between power and subjective ratings for each frequency band. The red color indicates positive correlation, whereas the blue indicates negative correlation.

4. CONCLUSION

In this study we investigated the way perceived QoE from 2D and 3D videos affects the brain. Initially we analysed the self-assessed ratings in terms of overall perceived quality from sixteen subjects and we showed that the actual quality and the content affect perceived quality, as expected. Moreover, from the subjective-rating analysis, it was also shown that interactions between content and actual quality, as well as between rendering mode and actual quality also affect perceived quality. Regarding the EEG correlates, the study revealed frontal asymmetry in the alpha EEG band, according to which the alpha power in the right frontal lobe is positively correlated with perceived quality, indicating that this lobe is activated when perceived quality is low. Finally, right parietal lobe was found to be activated when perceived quality is high, which according to neurophysiology indicates that perceived high quality is related to positive emotional processes, whereas perceived low quality to negative ones.

REFERENCES

[1] H.-C. O. Li, J. Seo, K. Kham, and S. Lee, "Measurement of 3D visual fatigue using event-related potential (ERP): 3D oddball paradigm," in *Proc. 3DTV Conf.*, Istanbul, Turkey, 2008, pp. 213–216.

[2] D. Kim, Y. J. Jung, E. Kim, Y. M. Ro, and H. W. Park, "Human brain response to visual fatigue caused by stereoscopic depth perception," in *Proc. Int. Conf. Digital Signal Processing*, Corfu, Greece, 2011, pp. 1–5.

[3] 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. 2619–2629, 2012.

[4] ITU-R BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," International Telecommunication Union, August 2012.

[5] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, April 2008.

[6] J.A. Coan and J.J.B. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biological psychology*, vol. 67, no. 1, pp. 7–50, 2004.

[7] R.J. Davidson, "What does the prefrontal cortex do in affect: perspectives on frontal EEG asymmetry research," *Biological psychology*, vol. 67, no. 1, pp. 219–234, 2004.

[8] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, no. 4700, pp. 750–752, 1985.