Explicit and Implicit Measures in Video Quality Assessment

Maria Laura Mele[1,2,3], Damon Millar[1] and Christiaan Erik Rijnders[1]
[1]COGISEN Engineering Company, Rome, Italy
[2]Department of Philosophy, Social and Human Sciences and Education, University of Perugia, Perugia, Italy
[3]ECONA, Interuniversity Centre for Research on Cognitive Processing in Natural and Artificial Systems, Sapienza University of Rome, Rome, Italy

Keywords: Video Quality Assessment, Single Stimulus Assessment Methods, Psychophysiological Assessment of User Experience, UX.

Abstract: This work investigates the relation between subjective Video Quality Assessment (VQA) metrics and psychophysiological measures of human interaction assessment such as gaze tracking, electroencephalography and facial expression recognition. Subjective quality assessment is based on deliberate judgement attributions of perceived quality and processes that human perceivers are not consciously aware of. Traditional VQA methods ask participants to deliberately assign a quality score to videos in terms of the perceptual video quality. A methodology combining psychophysiological measures with traditional VQA methods is rarely used in the literature. This paper describes a model of video quality assessment which takes into account both explicit and implicit measures of subjective quality, by addressing two questions: (1) Do traditional video quality assessment methods correlate with unaware/implicit psychophysiological measures of quality perception assessment? (2) What can the main psychophysiological methods add to traditional video quality assessment? Findings show that (1) psychophysiological measures are able to measure differences of perceptual quality in compressed videos in terms of number of fixations and that (2) both VQA methods and psychophysiological assessment methods combined are able to provide additional information about cognitive and affective processes of attribution of the affective factors that underlie the attribution of quality.

1 BACKGROUND

Subjective Quality Assessment is a key methodology to evaluate humans' perceived image or video quality in the most reliable way because objective models of video quality assessment are still unable to perfectly model human perception. Accurate perceptual models have become important because video compression uses them to remove information that is not perceived. Subjective video quality assessment aims to evaluate the perceived quality of videos processed with different coding parameters or distortion levels. In particular, Video Quality Assessment (VQA) methods are intended to test video processing systems either under optimum conditions (in which human observers are asked to assign a quality score to video-frames) or under non-optimum video signal transmission or emission (in which participants are asked to assign an impairment value to videos-frames that can be processed or not) (International Telecommunication Union, 2008). In VQA methods, users' average quality opinions are usually called Mean Opinion Score (MOS).

Traditionally, two main methods are used for subjective VQA, (1) the Single Stimulus method and (2) Double Stimulus method. (1) The Single Stimulus method shows consecutive sequences of single videos to the observers, who assign a quality score to each video. (2) The Double Stimulus method requires observers to assess two simultaneously displayed versions of the same video. In Double Stimulus methods, the observers are not informed whether one of the two versions is unprocessed (International Telecommunication Union, 2008).

In the literature, the most used subjective VQA methods are based on standardized models recommended by the International Telecommunication Union (ITU) of Geneva (Switzerland), which proposes different direct scaling methods for subjective testing (International Telecommunication Union, 1999; International Telecommunication Union, 2002). The ITU VQA

Mele, M., Millar, D. and Rijnders, C.
Explicit and Implicit Measures in Video Quality Assessment.
DOI: 10.5220/0007396100380049
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 38-49
ISBN: 978-989-758-354-4

methods are widely used in video processing practice. Subjective VQA methods are able to investigate the quality of video signals when averaged over many subjects, however, they have known limitations to take into account. One of the main limitations of subjective VQA methods is that MOS of quality ratings are highly variable across subjects. Subjective MOS are the result of conscious and deliberate cognitive processes of judgement attribution, which is likely to be influenced by different subjective factors such as expectation, biases, or cognitive strategies. Moreover, subjective evaluations can be affected by motivational or emotional factors such as social desirability. As the ITU claims, these known limitations (International Telecommunication Union, 2002) suggest that it may be unwise to place too much weight on a single method (International Telecommunication Union, 2002). Thus, it may be appropriate to consider more complete approaches such as the use of a multiple methods approach (Engelke et al., 2017).

Implicit and not immediately deliberate processes underlying quality perception of visual information are not directly measured by ITU subjective methods. General studies on decision making and quality judgement agree that cognitive phenomena such as workload, distraction, or avoidance/approach, influence the way in which information is transferred and presented to a decision maker (Wickens, 2004; Smith and Mistry, 2009). Previous studies proposed alternative VQA methods based on implicit measurements of psychophysiological activity to quantify perception of video quality with humans. For example, some studies (Antons et al., 2014; Arnau-Gonzalez et al., 2017; Jia et al., 2018; Scholler et al., 2012) investigated the process of human quality assessment of spatially degraded videos by measuring brain activity and/or eye movement parameters. Results showed that the perception of degraded areas of video-frames is related to higher levels of attention with an increase of pupil diameter and a decrease in the proportion of electroencephalographic (EEG) alpha activity (8-12 Hz), which is related to higher cortical activity (Gruzelier, 2014). Moreover, some studies with right-handed subjects discovered that asymmetries in frontal alpha power (Hagemann et al., 2002) occur either when left-frontal alpha increases (usually as a psychophysiological reaction to positive experiences of information processing), or when right-frontal alpha increases (usually as a reaction to negative experiences of information processing). Handedness affects asymmetric hemispheric activation in the degree to which each hemisphere is engaged. Therefore, magnitude of alpha asymmetry may be different with left-handed subjects (Davidson, 1988).

Using EEG for VQA is particularly useful for studying unconscious changes of perceived quality in time, and it does not require deliberate actions by observers to give any explicit video quality rating (Moldovan et al., 2013). Although these psychophysiological measures have previously been studied individually, this paper attempts to comprehensively study all of them simultaneously.

Image and video quality research has studied the benefits of using eye tracking methodology to measure interest, attention and workload. However, using eye movement measures as a unique method to directly evaluate subjective video quality perception might be susceptible to a main problem: salience maps are robust to compression. In other words, the average gaze position cannot be used to measure quality because salience is invariant to video quality (Albanesi and Amadeo, 2011). Temporal components of gaze such as fixation duration and fixation number are strictly related to decision-making processes, which remains the same regardless of compression. At a higher level, decision-making is known to increase the demand on attention and working memory. To lower the working memory load when it is becoming too demanding, the cognitive system decreases the processing accuracy by increasing the number of fixations and decreasing the fixation duration to compensate for working memory (Orquin and Loose, 2013). Saliency and quality values seem to be both correlated with the fixation duration so that fixations on liked and salient items last longer than fixations on disliked and less-salient parts (Towal et al., 2013; Majaranta, 2011). Based on these findings, fixation quantity and duration seem to form two independent axes providing a measure of the interest and cognitive demand, as figure 1 shows.

Two main methodologies are used in the User Experience (UX) field to measure emotions, (1) self-reports and (2) psychophysiological methods. The self-report methods measure the subjectively and consciously experienced emotions during an interaction, whereas the psychophysiological methods measure continuous emotional reactions and changes in central, autonomic and somatic nervous system (McDuff, 2017). Emotions and affective valence metrics are typically used to evaluate the emotional impact of contextual factors, the content of videos under assessment, or

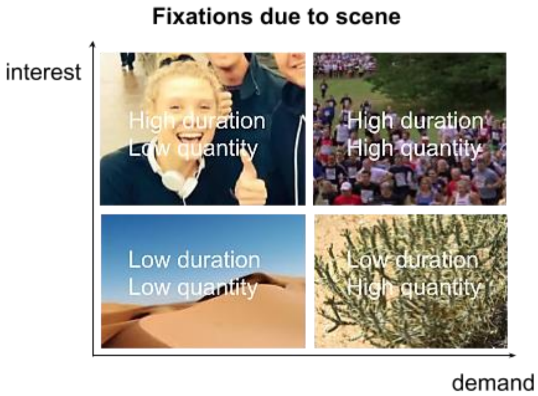

Figure 1: Visually complexity is related to both saliency and demand.

equipment and settings (De Moor et al., 2014; Ketyko et al., 2010; Reiter et al., 2012; Song et al., 2014). A recent work conducted by Msakni and Youssef in 2016 studied the impact of user affective valence/emotion and video content on users' experience of interaction, showing that (i) a bad/good evaluation seems not to be linked to a positive/negative emotion, (ii) visual content affects emotion after a subjective test (Msakni and Youssef, 2016). No study so far uses emotions and valence to evaluate whether differences in video processing affects emotions, in combination with other psychophysiological measures such as EEG or eye tracking.

A recent study by Tauscher and colleagues (Tauscher, 2017) compared user ratings, eye tracking, and EEG to investigate the differences between conscious responses and implicit perception. The work showed marked differences between fixations (eye tracking), consciously ratings of video quality, and neural responses (EEG data), concluding that each modality does not lead to a definitive conclusion if only considered in isolation. On one hand, user ratings methods differ from psychophysiological methods in that brain activity, eye movements or facial expressions are mostly controlled by unaware processes, so they can provide insights on the implicit processes underlying VQA. On the other hand, rating tasks may also involve aware and deliberate decision-making processes. Therefore, explicit measures miss out on implicit unconscious information.

Starting from the consideration that the cognitive processes required by video quality assessment for judging perceived quality are influenced by both implicit and explicit processes, this work focuses on two questions: (1) Do video quality assessment methods based on Mean Opinion Scores (MOS) relate with psychophysiological implicit measures of user experience such as visual attention, emotions, workload or engagement? (2) What additional information do psychophysiological measures give to traditional video quality assessment methods? The answer to these questions provides a model of video quality assessment that also takes into account implicit measures of subjective quality evaluation.

2 METHODOLOGY

The aim of this work is to investigate the relationship between aware (explicit) judgments of video quality and unaware (implicit) psychophysiological measures of video perception in a traditional Mean Opinion Score based Video Quality Assessment method.

The work applies a Single Stimulus VQA method to evaluate two different compression methods at two different Constant Rate Factors using a validated database that was assessed by the authors in previous studies (Mele et al., 2017a, Mele et al., 2017b; Mele et al., 2018). The Single Stimulus VQA method was arranged and administered by a biometric research software tool that is able to integrate and synchronize presentations, eye tracking, facial expression analysis, and electroencephalography (EEG) in one single platform.

2.1 Subjective Video Quality Assessment Methods

This study adopts a Single Stimulus method because it better complies with the home viewing conditions (International Telecommunication Union, 1999). In particular, the method used in this study is called Single Stimulus Continuous Quality Scale (SSCQS) (International Telecommunication Union, 1999), which is a hidden reference subjective VQA method displaying each test distorted video at a time and only once in one session. "Hidden reference" means that the reference high quality videos are randomly shown in the test as a control condition, without warning participants that that they are observing a high-quality video. At the beginning of each session, a brief training sequence of a few videos is introduced. Sequence presentation is randomized in order to ensure that the same video material is never presented twice one after the other. A typical assessment trial consists of an adaptation field, a stimulus field, and a post-exposure field. In the

SSCQS method, observers are asked to assign a quality value to each single video on a graded scale shown after each video ends. The quality scale that we used consisted of integers in the range 1-100. The scale is marked numerically on a slider and divided on into five equal portions, which are labelled with the following adjectives: “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”. The position of the slider is automatically reset to 1 after each evaluation ends.

The training trials appear before any test session without any noticeable interruption to the subjects. This procedure helps ensure the observers’ opinion is stabilized. The whole session should last no longer than fifteen minutes to avoid participants’ workload or learning effect. At the end of the test, Mean Opinion Scores (MOS) are calculated in an aggregated form. The quality scores collected during training trials should not be included in data analysis at the end of the test. The raw opinion scores are then converted to difference mean opinion score (DMOS) where the difference in scores between reference and distorted video is calculated (Sheikh et al., 2006).

2.2 Psychophysiology Methods

This study adopts three psychophysiology methods to evaluate the implicit components of subjective quality evaluation process: (1) facial expression recognition; (2) electroencephalography, and (3) eye tracking.

(1) *Facial Expression Recognition* is based on the fact that a highly significant combination of affective valence and emotional states is recognisable on our face. The movement of certain muscles of eyes, brows, lids, nostrils, and lips may provide information on the quality of the emotional response.

(2) *Electroencephalography* measures brain electrical signals over time. When neurons are activated, they produce a synaptic current that generates an electrical field over the scalp. This activity can be detected by EEG systems. EEG has been used in the UX fields to measure cognitive states during a task (Chai et al., 2014).

The most observed EEG rhythm in UX studies is the alpha wave, among five principal waves that can be distinguished with EEG according to their frequency range, i.e. delta (0.5 – 4 Hz); theta (4 - 8 Hz); alpha (8 - 12 Hz), beta (12 - 25Hz), and gamma (greater than 25Hz). Alpha activity can become suppressed/desynchronized when mental activity increases or when subjects become alert or drowsy (Pizzagalli, 2007). Alpha power increase is related with an increase in relaxation, and it is therefore inversely related to cortical activity (Allen et al., 2003). Frontal EEG asymmetry in alpha oscillations is considered as the most fundamental decision-making dimension employed by humans in terms of approach/withdraw processes (Schneirla, 1959). An alpha activity lateralized to the left hemisphere indexes tendencies to approach, whereas right alpha activity lateralization is related to withdraw from unexpected or affective stimuli (Coan and Allen, 2003). Alpha activity happens over the posterior head regions; therefore, the alpha frontal asymmetry can be calculated by the electrodes located at frontal scalp regions (usually, F3 and F4) and its amplitude (Brain symmetry index, BSI) is defined by the following ratio: (1)

$$\text{BSI} = (\text{left} - \text{right})/(\text{left} + \text{right}) \qquad (1)$$

Frontal alpha asymmetry (FAA) calculation (John, 1977) requires a preliminary pre-processing of the raw data which generally follows two steps: (i) epoch the data, i.e. break data into smaller temporal parts of up to 2 seconds; (ii) apply a frequency transformation to each epoch to determine frequencies; and (iii) the alpha frontal asymmetry index, generally computed as (2):

$$\text{FAA} = \log(\text{F4}) - \log(\text{F3}) \qquad (2)$$

where the difference between the alpha EEG power right electrode (usually F4) and the alpha EEG power left (usually F3) is calculated (Allen et al., 2004).

(3) *Eye Tracking Methodology* is a set of methods and techniques usually based on a corneal reflection system that is able to detect and record gaze position and eye movements in a given visual field. Most of the eye trackers in the market combine near-infrared technology with a high-resolution optical sensor to measure Pupil Center Corneal Reflection (PCCR). In PCCR, an infrared light source is used to elicit a reflection of the cornea and the pupil, that is then captured by a high-resolution camera. This process allows image-processing algorithms to measure the point of gaze related to the eye. Eye movements consist of a combination of saccades and fixations, and they are defined by both space and time. A saccade is defined as the movement between two fixations. Fixation duration is the length of time (usually 100-500 milliseconds) in which subject gazes on an area and it indicates attention at a specific location of the stimulus. Fixation area can be mapped to specific x-y coordinates and it may

indicate where user is paying attention (Duchowski, 2007). As eye movement indicate many aspects of cognition such as problem solving, reasoning, or search strategies, psychological research (as such as applied research in UX and human factors) predominantly uses eye-tracking methodology to gain insights into cognitive processes behind human behaviours (Ball et al., 2003; Just and Carpenter, 1976).

2.3 Apparatus

The psychophysiological instruments described as follows were integrated and synchronized in a biometric research tool called iMotions, which integrates and synchronizes sensors in one software platform with some types of stimuli such as videos, websites, screen recordings, or surveys (www.imotions.com).

(1) *Facial Expression Recognition*. In this work the Affectiva Affdex technology is used, provided as an integrated module in iMotions. Affectiva is a facial expression algorithm that implements the Facial Action Coding System (FACS) developed by Paul Ekman to code 24 core facial Action Units (Ekman and Friesen, 1978) that humans do without any deliberate control of them. Affectiva is able to measure affective valence (positive, negative and neutral (-100 to 100) and the seven basic emotions (1-100) proposed by Ekman, i.e. anger, contempt, disgust, fear, joy, sadness, and surprise. In this work, valence and emotional values were computed with a threshold of 20% confidence, meaning that only facial expressions with at least a 20% probability of a human assessor rating the emotion equally to the Affdex algorithm should be accepted. Each rating has an accompanying confidence. The expressions falling outside the threshold were labelled as neutral or as a lack of facial expressions.

(2) *Electroencephalography*. In this study the 256Hz B-Alert X10 EEG Headset System (Advanced Brain Monitoring, CA, USA), 10-Channel Wireless EEG Headset was used. As described by the manufacturer, "B-Alert is a Bluetooth wireless system and sensor headset to record up to 9 channels of monopolar EEG (Fz, F3, F4, Cz, C3, C4, POz, P3, and P4) based on the 10-20 system (where electrodes are separated by 10%–20% of the total distance around the circumference of the head), plus one optional channel of ECG data. "B-Alert Cognitive States Analysis software can be additionally used to create a *Benchmark* file of a subject's EEG profile by administering a series of simple onscreen tests and storing the resulting session data as a permanent reference for future EEG recordings" (Advanced Brain Monitoring, 2016) (www.biopac.com). All EEG channels were referenced to the mean of the left and right mastoids. In this study, the contact impedance between electrodes and skin was kept to a value less than 40 kiloohm (kΩ). A conductive electrode cream (Kustomer Kinetics, CA, USA) was applied to each electrode, including the reference, after cleaning its surface with 70% isopropyl alcohol.

(3) *Eye Tracking*. In this work we used a USB screen-based device called EyeTech VT3 Mini, with 120 Hz sample rate, 0.5° accuracy, and a 20 x 5 cm headbox size.

The whole experiment was carried out using a MSI laptop computer 7RE Dominator Pro with a 7th Gen. Intel® Core™ i7 processor, running Windows 10 Pro, GeForce® GTX 1070 8GB GDDR5, 17.3", 120 Hz Refresh Rate, 17.3" built-in 4K LCD with 3840 X 2160 resolution.

2.4 Materials

In this work, we replicated a subjective VQA test previously conducted and validated in a previous study (Mele et al. 2018). The subjective test uses five high technical-complexity benchmark videos derived from a set of source videos, each one lasting a mean of 10 seconds. The source database used is commonly used in the VQA field. It reflects a diversity in content and was previously validated by the authors in different studies (Mele et al., 2018; Mele et al., 2017). The effects of video content on psychophysiological data have not been studied before this study. Videos were 426 x 224 landscape resolution in the uncompressed YUV4MPEG 4:2:0 format. Reference videos were pre-processed with a visually lossless Constant Rate Factor (CRF) value of 10, and then processed by both the H264 model and a saliency-based model using two CRFs values (CRF=21 and CRF=27).

2.5 Procedures

The Single Stimulus Continuous Quality Scale survey was administered to participants using iMotions. The test was conducted under an artificial constant dim light in a UX laboratory equipped as described in section 2.3. Before starting the test, participants were informed about the general aims of the test and invited to read and sign the consent form. Each trial was carried out in a comfortable chair positioned about 60 centimetres from the screen. Both chair and desk height were adjustable to participants' height and requirements. Participants' data was anonymized. All subjects received financial compensation for their participation.

After participants were equipped with the EEG headset, a preliminary configuration of the equipment was carried out by the following steps: (i) Impedance check of each electrode site of the EEG headset, which measures the resistance between the scalp and electrode in kΩ (lower values mean better conductivity between scalp and electrodes). (ii) Acquisition of benchmark data to create individualized EEG profiles, which allow cognitive states measured to be valid and accurate across all participants. Benchmarking session consists of three vigilance tasks, ie, a three-choice vigilance task; a visual psychomotor vigilance task, and an auditory psychomotor vigilance task. Benchmark data collection typically requires 9-10 minutes. (iii) A 16 points eye-tracking calibration, asking the subject to fixate on 16 targets moving from central to peripheral positions. The whole pre-setting phase required at least 15 minutes per participant.

Once all psychophysiological devices were calibrated, participants were asked to observe a sequence of 3 trial videos plus 25 experimental videos and rate the perceived quality of each on a slider marked from 1 to 100. The sequence composed by 25 experimental videos lasted about 10 minutes.

2.6 Subjects

Nineteen right handed participants completed the subjective test in July 2018, 52.63% male, mean age 30 years old, no expert viewers. Participants were recruited by an online recruitment survey and they were selected after a preliminary interview about visual acuity, colour blindness, and professional experience in the field of video systems. Three outliers were excluded from the EEG dataset, and four participants were excluded by the eye tracking dataset.

3 RESULTS

Results on the subjective VQA survey, electroencephalography, eye tracking, and facial expression recognition are described as follows.

3.1 Subjective Video Quality Assessment Survey

For each subject, the Mean Opinion Scores assigned to the reference videos (REF MOS = 75.72; H264 CRF 21 = 70.83; Saliency-based CRF 21 = 71.84; H264 CRF 27 = 55.79; Saliency-based CRF 27 = 58.25) (Figure 2) were used to calculate the Difference Mean Opinion Scores (DMOS) (DMOS H264 CRF 21 = 5.40; DMOS Saliency-based CRF 21 = 3.89; DMOS H264 CRF 27 = 18.50; DMOS Saliency-based CRF 27 = 16.59).

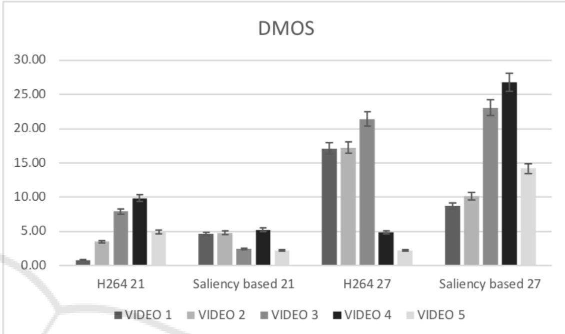

Figure 2: Difference Mean Opinion Scores (lower values mean higher quality scores) for compression types.

The repeated measures ANOVA on MOS shows a significant difference between reference videos and compressed videos (reference vs H264 CRF 21, $F(1,18) = 15.697$, $p = 0.001$; reference vs saliency-based CRF 21, $F(1,18) = 5.083$, $p = 0.037$; reference vs H264 CRF 27, $F(1,18) = 34.718$, $p = 0.000$; reference vs saliency-based CRF 27, $F(1,18) = 31.018$, $p = 0.000$), with lower MOS for compressed videos. No significant difference in the DMOS assigned to the H264 compressed method compared to the DMOS assigned to videos compressed with the saliency-based compression method was found for both CRF values (CRF 21, $F(1,18) = 1.870$, $p > 0.05$; CRF 27, $F(1,18) = 3.665$, $p > 0.05$).

3.2 Electroencephalography

The B-Alert EEG system provides an advanced automatic decontamination process for artefact removal, which minimizes the effects of muscular movements, eye movements, spikes, saturation, and excursions by a discrete wavelet transform (BIOPAC Systems, 2016). After pre-processing, the following scores are calculated by the B-Alert EEG system: (i) the engagement level (which is related to information-gathering, sustained attention, and visual scanning) which produces values from zero (low engagement) to one (high engagement), according to four classification levels (High engagement = 0.9, Low engagement = 0.6, Distraction = 0.3, Sleep onset = 0.1); (ii) the workload level, reflecting the level of working memory, problem solving and analytical reasoning during the trial. Workload values go from zero to one, where Boredom = up to 0.4, Optimal workload = 0.4-0.7, Stress and information overload = above 0.7 (iMotions, 2013); (iii) alpha frontal asymmetry, which is calculated as explained in section 2.4.

Engagement. For each video, the mean high engagement values (REF videos = 0.368; H264 CRF 21 videos = 0.386; Saliency-based CRF 21 videos =0.389; H264 CRF 27 videos = 0.378; Saliency-based CRF 27 videos = 0.363) and the mean low engagement values (REF videos = 0.302; H264 CRF 21 videos = 0.289; Saliency-based CRF 21 videos = 0.300; H264 CRF 27 videos = 0.049; Saliency-based CRF 27 videos = 0.305) were calculated (Figure 3).

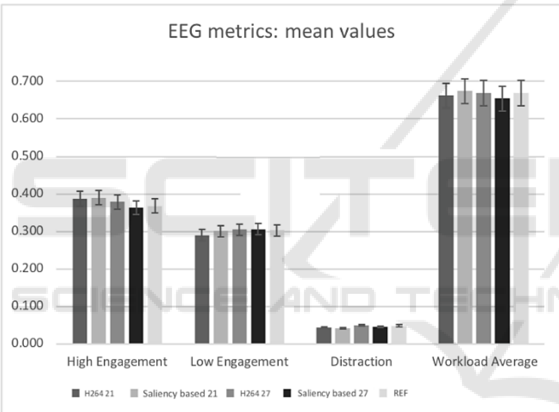

Figure 3: EEG values showing the levels of high engagement, low engagement, distraction, and workload for both compressed and reference videos.

Distraction. For each video, the mean distraction values were calculated (REF videos = 0.048; H264 CRF 21 videos = 0.042; Saliency-based CRF 21 videos = 0.041; H264 CRF 27 videos = 0.049; Saliency-based CRF 27 videos = 0.045) (Figure 3).

Workload. For each video, the mean workload values were calculated (REF videos = 0.669; H264 CRF 21 videos = 0.662; Saliency-based CRF 21 videos = 0.674; H264 CRF 27 videos = 0.669; Saliency-based CRF 27 videos = 0.654) (Figure 3).

Alpha Frontal Asymmetry. For each compression type, mean alpha asymmetry values (REF videos = 2.322; H264 CRF 21 videos = 2.278; Saliency-based CRF 21 videos = 2.342; H264 CRF 27 videos = 2.277; Saliency-based CRF 27 videos = 2.245) were calculated (figure 4).

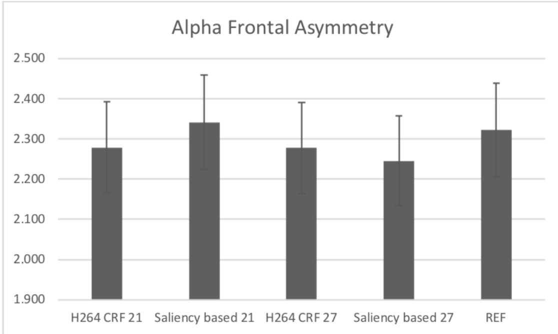

Figure 4: EEG values showing the levels of Frontal Alpha Asymmetry for both compressed and reference videos.

The repeated measures ANOVA on valence means (engagement, distraction, workload, alpha frontal asymmetry) show no significant difference between reference videos and compressed videos and between h264 and saliency based compressed videos ($p > 0.05$).

3.3 Eye Tracking

Fixation number and duration of fixation time were calculated for the test videos (Table 1) and the rating pages (Table 2).

Table 1: Mean fixation duration and Mean Fixation Time for compressed and reference videos.

TEST VIDEOS	Mean Fixation Number	Mean Fixation Time (ms)
REF	12.865	318.270
H264 CRF 21	13.217	308,27
Saliency-based CRF 21	12.625	317.954
H 264 CRF 27	12.995	312.770
Saliency-based CRF 27	15.960	335.929

Table 2: Mean fixation duration and Mean Fixation Time for the rating pages of compressed and reference videos.

RATING PAGES	Mean Fixation Number	Mean Fixation Time (ms)
REF	1.138	177.31
H264 CRF 21	1.22	175.60
Saliency-based CRF 21	1.081	180.23
H264 CRF 27	1.131	153.148
Saliency-based CRF 27	1.115	149.202

Fixation Number. The repeated measures ANOVA on eye-tracking metrics referring to test videos show a significant difference in fixation number between reference videos and H264 CRF 21 compressed videos ($F(1,14) = 152.962$, $p = 0.000$) and between reference videos and saliency based CRF 27 compressed videos ($F(1,14) = 5.111$, $p = 0.04$). No significant difference was found in fixation number between reference videos and both H264 CRF 27 videos (CRF 21, $F(1,14) = 0.165$, $p > 0.05$) and saliency based CRF 21 compressed videos (CRF 21, $F(1,14) = 0.602$, $p > 0.05$). A significant difference between the H264 compressed and saliency based compressed videos was found (CRF 21 = $F(1,14) = 98.368$, $p = 0.000$; CRF 27 $F(1,14) = 6.241$, $p > 0.026$)

Fixation Time. A significant difference in fixation duration was found between reference and CRF 27 compressed videos ($F(1,14) = 49.651$, $p = 0.000$). No significant difference in fixation duration was found between reference and CRF 21 compressed videos (H264 $F(1,14) = 2.426$, $p > 0.05$; saliency based = $F(1,14) = 0.001$, $p > 0.05$; CRF 27, $F(1,14) = 2.903$, $p > 0.05$). No significant difference between H264 compressed and saliency based compressed videos was found for fixation duration (CRF 21 = $F(1,14) = 0.395$, $p > 0.05$; CRF 27 = $F(1,14)= 2.654$, $p > 0.05$).

3.4 Facial Expression Recognition

The mean time percent for each emotion, calculated as the mean of 100*(count frames in which emotion appears/count frames in stimulus), was obtained for overall valence and basic emotions, as reported in Table 3.

Table 3: Mean time percent of both emotions and valence for compressed and reference videos.

	H264 CRF 21	Saliency-based CRF 21	H264 CRF 27	Saliency-based CRF 27
Positive	1.048	0.839	0.208	0.003
Negative	0.309	0.641	1.002	0.293
Neutral	78.991	75.296	76.564	75.888
Anger	0	0.003	0.01	0
Joy	0.414	0.579	0.282	0
Surprise	2.808	1.775	1.188	1.044
Sadness	0	0	0.032	0
Contempt	0.766	0.214	0.64	0.44
Fear	8.452	6.152	6.859	4.809
Disgust	0	0	0.458	0.003

The repeated measures ANOVA on valence means (positive, negative, and neutral) show no significant difference between reference videos and compressed videos and between h264 and saliency based compressed videos ($p > 0.05$).

3.5 Comparative Analyses

MOS vs Psychophysiological Measures. In test videos, a significant positive correlation was found between MOS and neutral affective valence (Pearson's $r = 0.420$, $p = 0.037$). No significant correlation was found between MOS scores and the other EEG metrics (engagement, distraction, workload) and between MOS scores and eye tracking values (number of fixations, fixation time). A significant positive correlation was also found between the duration of the test and alpha frontal symmetry mean scores (Pearson's $r = 0.500$, $p = 0.011$).

EEG Measures vs Eye Tracking Measures. Correlations between EEG values and eye tracking values were computed, showing no significant correlation among EEG values and eye tracking values for both test video stimuli and rating pages.

EEG Measures vs Facial Recognition Measures in Test Video Pages. Correlations between EEG data and affective values calculated in test videos showed a positive correlation between high engagement and both joy (Pearson's $r = 0.531$, $p = 0.006$) and positive valence (Pearson's $r = 0.559$, $p = 0.004$), a negative correlation between low engagement and both joy (Pearson's $r = -0.439$, $p = 0.028$) and positive valence (Pearson's $r = -0.460$, $p = 0.021$). Moreover, a negative correlation between contempt and workload was found (Pearson's $r = 0.552$, $p = 0.004$).

EEG Measures vs Facial Recognition Measures in Rating Pages. Correlations between EEG data and affective values calculated in rating pages showed a positive correlation between high engagement and fear (Pearson's $r = 0.466$, $p = 0.022$), a negative correlation between low engagement and both joy (Pearson's $r = -0.439$, $p = 0.002$), and a positive correlation between workload and sadness was found (Pearson's $r = -0.449$, $p = 0.028$).

Eye Tracking Data vs Facial Recognition Measures in Test Videos. Correlations between eye tracking data in video pages and the related affective values were computed, showing no correlation between fixation number and duration, and between valence and emotions.

Eye Tracking Measures vs Facial Expression Measures in Rating Pages. Correlations between eye tracking data in rating pages and the related facial expression values showed a positive correlation between fixation number and positive valence (Pearson's $r = 0.584$, $p = 0.002$) and fixation number and surprise (Pearson's $r = 0.622$, $p = 0.001$). Similar correlations were found between fixation duration in rating pages and (i) positive valence ($r = 529$, $p = 0.007$), (ii) negative valence ($r = 562$, $p = 0.003$), (iii) surprise ($r = 0.761$, $p = 0.000$), and (iv) contempt ($r = 0.426$, $p = 0.008$). Moreover, a significant negative correlation between positive valence and the most fixed rating AOIs was found (Spearman's $r = 0.414$, $p = 0.04$).

4 DISCUSSION

The current study aimed to understand (i) in which measure current video quality assessment methods relate with psychophysiological measures, and (ii) whether psychophysiological measures add to traditional video quality assessment methods new valuable information about humans' experience of interaction.

(i) Which psychophysiological measure correlate with subjective video quality assessment measures? As expected, explicit subjective ratings showed a significant difference between reference and compressed videos. Observers assigned qualitatively higher scores to the test reference videos than the compressed ones, meaning that the observers noticed quality distortions in videos compressed with both methods.

Differences in the effects of video quality observation were also investigated among the psychophysiological values measured during the VQA test for reference and compressed videos, i.e. electroencephalography, eye tracking and facial expression recognition. A significant difference between compressed and reference videos was found only for eye tracking measures too. EEG and facial expression recognition measure did not result able to detect any significant difference between reference videos and compressed videos. Findings show significantly higher fixation number and fixation duration on compressed videos than on reference videos, thus replicating the results obtained for the explicit VQA method. In particular, fixation duration decreases for videos compressed at high compression levels, whereas fixation number is significantly higher in compressed videos than in reference videos. Findings on gaze shows then that fixations are sensitive to differences in video quality as such explicit VQA subjective ratings. Contrary to VQA, fixation measures seem to be also able to selectively indicate which of the compressed video types and compression levels significantly affect subjective quality perception. These results confirm that higher fixation number indicates an increase of working memory load, and that fixation time increases as mental load increases.

(ii) Do psychophysiological measures add to traditional video quality assessment methods new valuable information about human perception? Findings on electroencephalography values showed that frontal alpha asymmetry increases as the test progresses, thus indicating a downturn in the observers' approach motivation to interact with the system as the duration of the test increases. However, less motivation does not indicate a positive/negative affective valence, since a decrease in approach motivation does not correlate to valence. During the observation of test videos, neutral valence was significantly present than both positive and negative valence independently from compression level, confirming that a bad/good evaluation of video quality seems to be not linked to a positive/negative experience of interaction (Msakni and Youssef, 2016).

High engagement values measured through the EEG correlated with negative emotions of fear when rating pages were presented, suggesting that participants were afraid of giving a wrong answer while assigning scores to videos (even if they were previously warned that there were not correct answers to the test). High engagement correlated also with positive emotions of joy while observing videos. At the same time, engagement did not correlate with subjective scores of video quality. Therefore, the attention allocation processes behind perceivers' engagement are not linked to the explicit evaluation of video quality, but they might be related to video content.

Workload values measured through the EEG correlated with sadness in both rating pages and video pages. The more the perceived amount of work increases, the more negative emotions of sadness increase for participants. However, this did not affect the test since negative valence did not significantly occur during the test.

Finally, eye fixation duration in rating pages positively correlated with both positive and negative valence values and emotions of surprise, whereas eye fixation number positively correlated only with positive valence values, surprise and contempt. These findings suggest that emotions of surprise and affective non-neutral valence might indicate qualitative information of observers' experience of assigning a quality score. However, these results did not occur for video pages, showing no changes in emotions and valence for visual perception of videos but it only happens when users are involved in the rating task.

5 CONCLUSIONS

This work was able to (1) investigate in which way the logic behind one of the most used video quality assessment method, i.e. the Single Stimulus Continuous Quality Scale (SSCQS), relates with unaware psychophysiological measures of quality perception, and (2) understand what the main psychophysiological methods can add to traditional video quality assessment methods. Gaze tracking measures, electroencephalography and emotion recognition through facial expression recognition were used to measure implicit components of users' interaction with videos during a SSCQS-based survey. Reference videos and videos compressed with two different compression methods were used.

Main findings showed that gaze fixation measures are able to predict differences in perceived quality of video compression during VQA. Fixation number increases and fixation duration decreases as compression levels increment. Contrary to explicit VQA methods, fixation measures were more sensitive than explicit VQA methods to difference of quality in that eye tracking measures were able to better discriminate differences among compression levels.

EEG measures revealed that observers' approach motivation decreases as the duration of the video quality rating test increases. As high workload values increased, negative emotions of sadness increased too in both video and VQA ratings, showing a relation between workload and stress-related negative emotions. However, perceived workload was not significantly high during the entire test. Therefore, workload was not strong enough to affect overall participants' emotional valence, which was overall neutral during the whole test.

Emotions of fear appeared only during the rating task together with high engagement EEG values. Contrarily, during video display, high engagement EEG values related to positive emotions of joy but not to subjective scores of quality. Therefore, positive valence factors that are not linked to video compression (likely video content) increased participants' sustained attention.

In conclusion, psychophysiological measures do not fully account for quantitative explicitly perceived differences in video compression. Only gaze tracking measures predict VQA scores. However, measures of affective valence, basic emotions, engagement, workload and frontal asymmetry are able to add additional qualitative information of video quality. Future works will extend the implications of this study to other sensory stimuli such as audio or multimedia content, in order to study whether and how implicit processes affect subjective signal quality assessment processes.

REFERENCES

Albanesi, M. G., & Amadeo, R. 2011. Impact of Fixation Time on Subjective Video Quality Metric: a New Proposal for Lossy Compression Impairment Assessment. *World Academy of Science, Engineering, and Technology*, 59, 1604-1611, p. 1334).

Allen, J. J. B., Coan, J. A., & Nazarian, M. 2003. What's the difference? Issues and assumptions in the use of difference scores and other metrics of anterior brain asymmetry in emotion. Submitted to the special issue of *Biological Psychology*.

Allen, J. J., Coan, J. A., & Nazarian, M. 2004. Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biological psychology*, 67(1-2), 183-218.

Antons, J. N., Arndt, S., Schleicher, R., & Möller, S. 2014. *Brain activity correlates of quality of experience. In Quality of experience* (pp. 109-119). Springer, Cham

Arnau-Gonzlez, P., Arevalillo-Herrez, M., & Ramzan, N. 2017. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals, *Neurocomputing*, vol. 244, pp. 81 – 89.

Ball, L. J., Lucas, E. J., Miles, J. N. V., & Gale, A. G. 2003. Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, 56A, 1053-1077;

Billeci, L., Sicca, F., Maharatna, K., Apicella, F., Narzisi, A., Campatelli, G., Calderoni, S., Pioggia, G., Muratori, F., 2013. On the application of quantitative EEG for characterizing autistic brain: a systematic review, *Frontiers in Human Neuroscience*, Vol. 7, No. 442, pp. 1-15.

BIOPAC Systems, Inc. Manufactured by Advanced Brain Monitoring, 2016, Inc.B-Alert with AcqKnowledge Quick Guide Benchmark Acquisition and Cognitive States Analysis.

Chai, J., Ge, Y., Liu, Y., Li, W., Zhou, L., Yao, L., & Sun, X. 2014. Application of frontal EEG asymmetry to user experience research. In *International conference on engineering psychology and cognitive ergonomics* (pp. 234-243). Springer, Cham.

Coan, J. A., & Allen, J. J. 2003. Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysiolog*y, 40(1), 106-114.

Coan, J. A., & Allen, J. J. 2004. Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological psychology*, 67(1-2), 7-50.

Davidson, R. J. 1988. EEG measures of cerebral asymmetry: Conceptual and methodological issues. *International journal of neuroscience*, 39(1-2), 71-89.

Davidson, R. J. 2004. What Does the Prefrontal Cortex 'Do' in Affect: Perspectives on Frontal EEG

Asymmetry Research, *Biological Psychology*, 67, 219-33

De Moor, K., Mazza, F., Hupont, I., Quintero, M. R., Mäki, T., & Varela, M. 2014. Chamber QoE: a multi-instrumental approach to explore affective aspects in relation to quality of experience. In *Human Vision and Electronic Imaging XIX* (Vol. 9014, p. 90140U). International Society for Optics and Photonics.

Duchowski, A. T. 2007. *Eye tracking methodology. Theory and practice*, 328.

Ekman, P., Friesen, W.V., 1978. Manual for the facial action coding system. *Consulting Psychologists Press*.

iMotions,A.S. 2013. EEG iMotions Biometric Research Platform, from https://imotions.com/eeg/.

Engelke, U., Darcy, D. P., Mulliken, G. H., Bosse, S., Martini, M. G., Arndt, S., ... & Brunnström, K. 2017. Psychophysiology-based QoE assessment: A survey. *IEEE Journal of Selected Topics in Signal Processing*, 11(1), 6-21.

Gruzelier, J. H. 2014. EEG-neurofeedback for optimising performance. I: a review of cognitive and affective outcome in healthy participants. *Neuroscience & Biobehavioral Reviews*, 44, 124-141

Hagemann, D., Naumann, E., Thayer, J. F., & Bartussek, D. 2002. Does resting electroencephalograph asymmetry reflect a trait? An application of latent state-trait theory. *Journal of personality and social psychology*, 82(4), 619.

iMotions, A. S. 2015. Affectiva iMotions Biometric Research Platform. from https://imotions.com/ affectiva.

International Telecommunication Union, 1999. *ITU-T Recommendation P.910, Subjective video quality assessment methods for multimedia applications*, International Telecommunication Union. Geneva, Switzerland.

International Telecommunication Union, 2002. *ITU-R Recommendation BT.500-11, Methodology for the subjective assessment of the quality of television pictures*. Geneva, Switzerland.

International Telecommunication Union, 2008. *ITU: New definitions for inclusion in Recommendation ITU-T P.10/G.100, ITU-T Rec. P.10/G.100.* Geneva, Switzerland.

Jia, L., Tu, Y., Wang, L., Zhong, X., & Wang, Y. (2018). Study of image quality using event-related potentials and eye tracking measurement. *Journal of the Society for Information Display*.

John, E.R. 1977. Neurometrics: clinical applications of quantitative electrophysiology, in: E.R. John, R.W. Thatcher, (Eds), *Functional neuroscience*: vol.2, Lawrence Erlbaum Associates, Hillsdale, NY, pp. 246-251.

Just, M. A., & Carpenter, P. A. 1976. Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441-480.

Ketyko, I., De Moor, K., Joseph, W., Martens, L., & De Marez, L. 2010. Performing QoE-measurements in an actual 3G network. *In Broadband Multimedia Systems and Broadcasting (BMSB), 2010 IEEE International Symposium on* (pp. 1-6). IEEE.

Majaranta, P. (Ed.). 2011. Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: *Advances in Assistive Technologie*s. IGI Global.

Mele, M. L., Millar, D., & Rijnders, C. E. 2017a. The Web-based Subjective Quality Assessment of an Adaptive Image Compression Plug-in. In *VISIGRAPP (2: HUCAPP)* (pp. 133-137);

Mele, M. L., Millar, D., & Rijnders, C. E. 2017b. Using Spatio-Temporal Saliency to Predict Subjective Video Quality: A New High-Speed Objective Assessment Metric. In *International Conference on Human-Computer Interaction* (pp. 353-368). Springer, Cham.

Mele, M. L., Millar, D., & Rijnders, C. E. 2018. Sencogi Spatio-Temporal Saliency: A New Metric for Predicting Subjective Video Quality on Mobile Devices. In *International Conference on Human-Computer Interaction* (pp. 552-564). Springer, Cham.

McDuff, D. 2017. "New Methods for Measuring Advertising Efficacy". *Digital Advertising: Theory and Research*.

Song, W., Tjondronegoro, D., & Himawan, I. 2014. Acceptability-based QoE management for user-centric mobile video delivery: A field study evaluation. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 267-276). ACM.

Moldovan, A. N., Ghergulescu, I., Weibelzahl, S., & Muntean, C. H. 2013. User-centered EEG-based multimedia quality assessment. In *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on* (pp. 1-8). IEEE.

Msakni, H. G., & Youssef, H. 2016. Impact of user emotion and video content on video Quality of Experience. In *Proc. 5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)* (pp. 97-101)

Orquin, J. L., & Loose, S. M. 2013. Attention and choice: A review on eye movements in decision making. *Acta psychologica*, 144(1), 190-206.

Pizzagalli, D. 2007, Electroencephalography and high density electrophysiological source localization, in: J.T., Caccioppo, L.G., Tassinary, G. Berntson, (Eds), *Handbook of psychophysiology*, Cambridge University Press, Cambridge, UK, pp. 56-84.

Reiter, U., & De Moor, K.., 2012. Content categorization based on implicit and explicit user feedback: combining self-reports with EEG emotional state analysis. *In Quality of multimedia experience (QoMEX), 2012 fourth international workshop on* (pp. 266-271), IEEE.

Schneirla, T. 1959. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. Jones (Ed.), *Nebraska Symposium on Motivation* (pp. 1-42). Lincoln, NE: University of Nebraska Press.

Scholler, S., Bosse, S., Treder, M. S., Blankertz, B., Curio, G., Muller, K.-R., & Wiegand, T. 2012. Toward a direct measure of video quality perception using EEG, IEEE Trans. *Image Process*., vol. 21, no. 5, pp. 2619–2629.

Sheikh, H. R., Sabir, M. F., & Bovik, A. C. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11), 3440-3451.
Smith, K. T., Mistry, B. 2009. Predictive Operational Performance (PrOPer) Model. *Contemporary Ergonomics 2009 Proceedings of the International Conference on Contemporary Ergonomics*.
Tauscher, J. P., Mustafa, M., & Magnor, M. 2017. Comparative analysis of three different modalities for perception of artifacts in videos. *ACM Transactions on Applied Perception (TAP)*, 14(4), 22.
Towal, R. B., Mormann, M., & Koch, C. 2013. Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, 110(40), E3858-E3867.
Wickens, C., Lee, J., Liu, Y., Becker., S. E. G., 2004. An *Introduction to Human Factors Engineering*. Upper Saddle River, NJ: Pearson/Prentice Hall.
Winkler S. & Mohandas, P. 2008. The evolution of video quality measurements: from PSNR to hybrid metrics, *IEEE Transactions on Broadcasting* 54, pp. 660–668.