

# Image Quality in Augmented Binocular Vision: QoE Approximations with QoS and EEG

Patrick Seeling<sup>1\*</sup>

RESEARCH ARTICLE

Received 10 May 2016; accepted after revision 14 August 2017

## Abstract

Augmented reality (AR) applications become increasingly popular; however, little is known about how multimedia consumption interplays with the Quality of Experience (QoE) in these settings. We experimentally evaluate binocular vision augmentation with optical see-through devices by contrasting QoE, expressed by Mean Opinion Scores (MOS), with a ground truth reference data set. We find that the QoE in AR settings (i) is higher for small media impairments and lower for high impairment levels compared to opaque settings, (ii) exhibits an emerging quantifiable relationship with the QoE in traditional setups, (iii) can be approached with common objective image quality metrics as Quality of Service (QoS) factors, and (iv) exhibits a relationship between the Low Gamma frequency band levels determined with consumer-grade electroencephalograms (EEG) and image quality levels as rated by the participants.

## Keywords

Quality of Experience, Quality of Service, Augmented Reality, image quality, EEG

## 1 Introduction

A significant push for the utilization of Augmented Reality (AR) has been enabled through the developments in display technologies and smaller-scale computing. AR devices that are performing in a heads-up-display (HUD) or Head-Mounted Display (HMD) manner are increasingly targeting the military, industrial, and consumer application spaces alike, indicating future broad adaptation of the technology. While these types of devices are available in a broad variety of implementations (see, e.g., [1] for an overview of different types), a slow convergence of system types has begun, especially in the consumer space. In turn, a variety of commercial off-the-shelf (COTS) see-through wearable devices have emerged, which are the result of continuous developments for more than a decade and similar to the one presented in, e.g., [2]. Industry predictions indicate that future mobile networking scenarios will include a significant portion of multimedia consumed by users on-the-go [3], with single or binocular vision augmenting displays having the potential to be used in a considerable amount of future scenarios.

The evaluation of content in AR scenarios with optical see-through type devices has, to a large degree, focused on the assessment of the device operator performance in an holistic approach and [4] highlights persisting issues for the various system types. Evaluations performed additionally target the user-perception of augmentation for daily life scenarios, such as in [5], or how to limit the amount of additional information, as in, e.g., [6]. Perceptual evaluations oftentimes consider the segmentation of virtualized/augmented items, such as in, e.g., [7]. The estimation of a user's perception has attracted initial research, for example focusing on depth perception issues, e.g., in [8], where immersive displays were used. Similarly, contrast and color perception issues were evaluated in [9] with mixed results. Furthermore, it was found that binocular optical see-through HMD devices resulted in a negative impact on the field of view, which, in turn, could have a negative impact on the operator performance, see, e.g., [10]. Single view augmenting devices, on the other hand, did not exhibit this impact. A more recent overview of issues related to the perception in AR scenarios that focuses on the display modalities is provided

<sup>1</sup> Department of Computer Science,  
College of Science and Engineering,  
Central Michigan University,  
Mount Pleasant, MI 48859, USA

\* Corresponding author, e-mail: [patrick.seeling@cmich.edu](mailto:patrick.seeling@cmich.edu)

in [4]. While these significant research efforts target the overall device operator performance and calibration issues, little research to date considers the display of multimedia content for this type of device and the resulting experience that users derive from the device utilization.

To efficiently deliver multimedia data, such as video, common compression approaches target the optimization of the trade-off between amounts of data (bandwidth) required for content delivery and quality thereof, see, e.g., [11]. The manner in which users experience the media consumption in traditional settings has gathered significant research efforts over the past decade, moving from a network-centric determination of Quality of Service (QoS) to the user experience, referred to as Quality of Experience (QoE). The QoE has been principally defined as “[...] the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state” [12]. The subjective origins of experiences increase the complexity of evaluation, see, e.g., [13], commonly performed using the Experience Sampling Method (ESM), and is captured using Likert-scales to derive Mean Opinion Scores (MOS), see, e.g., [14, 15]. As the continuous involvements of users are not feasible for large-scale evaluations of media consumption experiences, relationships of the underlying objectively determined QoS metrics and the resulting QoE were recently combined into the exponential interdependency of QoE and QoS hypothesis (IQX) in [16] and further extended in [17]. Both approaches assume an exponential or logarithmic relationship between QoE and QoS based on the notion that the underlying levels of a stimulus change its perception, derived from the Weber-Fechner law of “just noticeable differences.”

An additional challenge for the display of any type of content using optical see-through devices stems from the overlay of the displayed content with that of reality. The outlined challenges for device operators in [18] for interactions, for example, are indicative of additional cognitive strains on the operator in specific situations. These could be explained through the dynamic interplay of real world background textures or color schemes and those provided by the optical see-through device. In turn, additional psycho-physiological aspects are commonly assumed to play a significant role in the evaluation of multimedia experiences. Video quality, for example, was found in prior research to have a correlation to cognitive load in [19]. Cognitive load theory has been applied to human learning processes in the past, see, e.g., [20]. The measurement of cognitive load during experimental processes, such as in, e.g., [21, 22], can thus be regarded as one additional foundational impact factor to relate quality impairments and their impact on user experience. The cognitive composition in augmented reality settings has attracted initial interests for modeling the

visual system, such as described in [23], where long-term and short-term memory processes are combined to identify objects in real-world contexts. Similarly, the authors of [24] are investigating the applicability of their perception model in terms of usability considerations. With the recent advances in Brain-Computer Interfaces (BCI), a direct measurement of brain activity and analysis through Electroencephalography is an interesting avenue. Earlier findings, such as in [25] indicate that the gamma frequency band might be of interest in the evaluation of visual stimuli, with a connection to memory processing in, e.g., [26]. Given the affective influence on the overall QoE as subjective metric, a BCI-based prediction of emotional state was presented in [27].

In this article, we describe an approach to the measurement of the Quality of Experience in augmented binocular vision scenarios, which is based on our previous works originally presented in [28] and [29].

We extend on these prior efforts by introducing a first in-depth evaluation of the relationship between electroencephalogram (EEG) readings and image qualities, measured objectively as QoS and indicated subjectively as QoE. The remainder of this article is structured as follows: In the subsequent Section 2, we describe the overall setup at the heart of our investigation. Next, we present a comparison of Mean Opinion Scores as expressed QoE with a popular database in Section 3. We continue with a comparative approximation of the QoE from objective image quality metrics as QoS indicators in Section 4 and provide a relationship between EEG measurements and QoS/QoE in Section 5. We discuss the results and implications in Section 6 before concluding in Section 7.

## 2 Experimental Setup and Metrics

In this section, we describe the overall approach to the determination of the QoE employing a binocular optical see-through device, following the initial guidelines presented for this evaluation scenario, see [29]. Most of the experimental configurations are shared with a traditional opaque display setting, similar to the overall descriptions found in, e.g., [30, 31]. Overall considerations for human subject experimentation and initial instruction are similarly shared.<sup>1</sup> The overall population of the participating subject group is comprised of 3 females and 9 males. All participants were young adults under 40 at the time of their experimental session [32], with an average age for the participants of 26.2 years (SD=5.31 years). The group was divided into 9 males and 3 females. Five subjects wore vision-correcting glasses or contact lenses for the duration of the experiment, while seven subjects were reporting no regular vision correction requirements.

---

<sup>1</sup> We note that the latest approval for this research was obtained on 02/18/2015 from the Institutional Review Board at Central Michigan University.

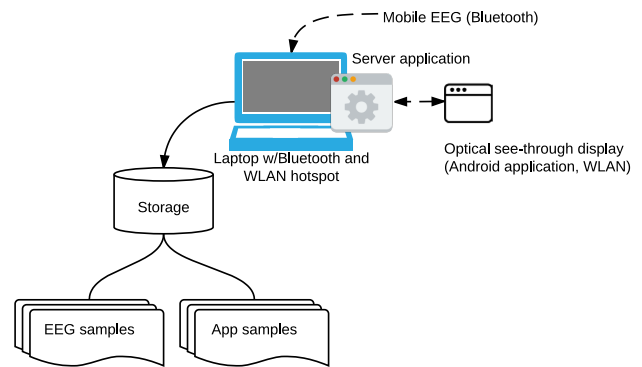
## 2.1 System and Devices Employed

While several vision augmenting device types exist to date, two main device types have emerged in the consumer space that employ an optical see-through approach, namely (i) single vision and (ii) binocular vision augmenting devices. The employed experimental configuration relies on the binocular Epson Moverio BT-100 mobile viewer, which is comprised of a head-mounted display unit and a central processing unit (utilizing the Android operating system) connected via wires. The display unit has a resolution of 960×540 pixels with Light Emitting Diode (LED) light sources and a 23 degree field of view. 24-bit color reproduction is available at 60 Hz and the built-in LED light intensity is set to maximum for highest contrast against different backgrounds. A generalized level of about 70 % transparency is realized for the display, where the light hits a half-mirror layer in a light guide material. The real-world background for this study was a whiteboard with approximately 50 lux of directional ambient light (from the right, others around 5 lux).

Measuring of cognitive load levels through electroencephalography (EEG) was performed in the experimental setting using the NeuroSky Mindwave Mobile headset, a consumer-grade wearable EEG piece of equipment. Similar single dry electrode devices were employed in prior research studies, see, e.g., [33, 34]. This type of device does not require an extensive setup procedure nor an expert for calibration and operation, but measurements employing this device type are limited to the FP1 region of the forehead. The typical configuration of both wearable devices as worn by subjects is illustrated in Fig. 1. Both head-mounted devices were connected to a laptop computer with a mobile hotspot for WLAN and Bluetooth connectivity. Fig. 2 illustrates the overall measurement environment. While the mobile EEG measurement device connects to the laptop using Bluetooth, the Android-powered display connects employing a laptop-originated wireless LAN hotspot. We chose this experimental configuration, as the communication delays for short-range networking are small and do not interfere with the overall setup. In turn, the time-stamping on the laptop can be performed for the EEG and display device measurements in a straightforward manner and allows for basic synchronization within COTS device accuracies. The interaction with the mobile Android application installed on the optical see-through display device is performed using a websocket server that executes on the laptop and a client connection from the custom viewer application. The server side controls the display of media by sending the image to be displayed to the mobile application and retrieving the user-selected rating.



**Fig. 1** Example configuration as worn by participating human subjects: consumer-grade EEG device (NeuroSky Mindwave Mobile headset) and binocular optical see-through vision augmentation device (Epson Moverio BT-100 mobile viewer).



**Fig. 2** Schematic view of the overall evaluation system and main components: laptop computer, mobile EEG device, and optical see-through device.

## 2.2 Experimentation

A significant amount of databases containing images and human subject evaluations exist to date, see, e.g., [35] for an overview. We selected the Tampere Image Database from 2013 (TID2013), as it is widely used and contains a significant number of image impairment variations and user ratings to provide a ground truth for comparisons, see [36]. We employ a subset of the images found in the database and limit the impairments to those that are results of JPEG encoding, due to time constraints for each individual subject's exposure to the experimental conditions. The selected images are illustrated in Fig. 3.



**Fig. 3** Images used for the perceptual quality evaluations in our pilot study (numbers 4, 7, 14, 15, 18, 21, and 23 from the Tampere Image Database, left to right) employing five different ranges of impairments from JPEG compression next to the original.

These images contain different content types and textures as well as foreground-background dynamics and, hence, cover a broad range of potential real-world content possibilities. The experimental flow transitions from initial explicit consent over a short questionnaire for age, gender, and visual impairment, which is followed by instruction of the overall experimental configuration and device usage. In our experimental setup, single stimulus presentations were employed with the absolute category rating with hidden reference approach (ACR-HR), where the subjects rate all, including the original media, on a five-point Likert scale. The participants viewed the individual media for 10 seconds, which was followed by an unconstrained voting time period to mitigate the impacts that the interface with unknown devices and potentially unfamiliar user interface design patterns might have on the individual participant's rating process.

### 2.3 Metrics

Let  $I_i^l$  denote the source images used from the TID2013 data set, with images  $i$ ,  $i \in \{4, 7, 14, 15, 18, 21, 23\}$  and impairment levels  $l$ ,  $l \in \{0, \dots, 5\}$ , whereby  $l = 5$  denotes the lowest quality (highest compression) and  $l = 0$  denotes the highest quality (uncompressed source image). Furthermore, we denote the average quality ratings in terms of MOS provided from the data set as  $\bar{M}_i^l$ . We note that the original ratings are provided on a 10-point scale ranging from 0 to 9; they were adjusted to a comparable 5-point scale as

$$\bar{M}_i^l = 4 \cdot \frac{\text{TID2013-MOS}(I_i^l)}{9} + 1. \quad (1)$$

Similarly, we denote the subject-selected image quality levels in the augmented binocular vision experiments for user  $j$  as  $q_{i,j}^l$ , noting that users were not able to choose an uncompressed quality level, i.e.,  $q_{i,j}^l \in \{1, \dots, 5\}$ . The resulting mean opinion score is subsequently denoted as  $\bar{Q}_i^l$  for image  $i$  at quality level  $l$ . As the overall setup includes the determination of the uncompressed image quality, we determine the differential MOS as  $\bar{Q}_i^l = \min(5, \bar{Q}_i^l - \bar{Q}_i^0 + 5)$ ,  $l = 1, \dots, 5$ . We denote the average opinion scores by quality level as

$$\bar{Q}^l = \frac{1}{\sum_i 1} \sum_i \bar{Q}_i^l \quad (2)$$

for the augmented vision scenario and  $\bar{M}^l$ ,  $\bar{Q}^l$  synonymously. Similar to the previous quality-related notations, let  $\bar{E}_{i,j}^l(b, t)$  denote the average captured EEG levels of subject  $j$  for image  $i$  for different common EEG frequency bands  $b$  over a specific period of time  $t$ ,  $t \in \{t_{\min}, \dots, t_{\max}\}$ . As indicated in the experimental design description, a subject views the image for a prescribed time of 10 s, such that we consider  $t_{\min} = 1$  s and  $t_{\max} = 10$  s. We note that aggregations on the EEG values are performed in a manner similar to those for MOS.

### 3 Mean Opinion Scores

We initially focus on the comparison of the MOS that were derived from the user ratings in the augmented vision configurations with those presented in the TID2013 data set. Initially, we compare the different MOS ratings for the selected images without any differential adjustments. We note that the impairment level  $l$  presented to the participants and their selection of quality  $q_{i,j}^l$  had an overall Pearson correlation coefficient average of 0.824 (SD=0.07). This indicates that on average, individual users' ratings exhibit a fairly high correlation to the presented impairment levels.

We illustrate the mean opinion scores in Fig. 4a for the different impairment levels and compare them to the scores reported for the TID2013 database. We initially observe that the MOS ratings  $\bar{Q}_i^l$  for the different images fall into a general band that starts in the upper ranges of opinion scores (around 4.4 on average) when not considering any impairments. Afterwards, the ratings hold somewhat steady for the lower impairment levels of one and two, only to fall off rather exponentially with increasing impairments. The exception here is image 14, which exhibits a lower score and immediately notable decline. These trends are to be expected, but sharper in their decline, when comparing these results with those reported from the TID2013 data set, which are illustrated in Fig. 4b. For the opaque scenario, we observe that the ranges of AR-MOS reported are spread further, which can be attributed to the numbers of ratings present in each data set and subsequent variability. Interestingly, we note a similar, but less pronounced, behavior for the image 14 in the TID2013 data set, which, in turn, can be attributed to the content features of the image. We perform a direct comparison of images and averages in Fig. 4c, whereby we subtract the AR-MOS from the ATID-MOS, i.e., we determine  $\bar{M}_i^l - \bar{Q}_i^l$ . We immediately observe an inverse behavior in comparison to the individual MOS obtained in either setting, with an overall shift in the sign in the medium impairment ranges. Specifically, we note that for low impairment levels,  $\bar{Q}_i^l$  is higher than  $\bar{M}_i^l$  in all cases. Even more interesting, the difference is almost steady for the impairment levels  $l$  one and two, and only afterwards decreases almost linearly until the TID2013 MOS  $\bar{M}_i^l$  outperform those determined for augmented vision  $\bar{Q}_i^l$ . This QoE difference effect for augmented versus opaque vision was anecdotally reported in [29], and can be regarded as the difference stemming from opaque and augmented vision presentation modes.

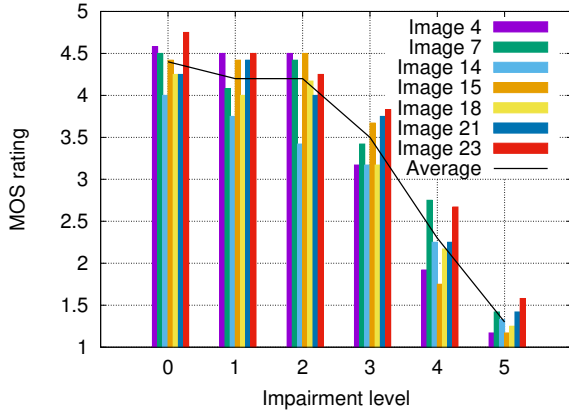
Subsequently, we evaluate the relationship between the image quality ratings obtained for both presentation modes in Fig. 5.

We observe that the functional relationship of augmented MOS and reported opaque presentation  $\bar{Q}_i^l = f(\bar{M}_i^l)$  exhibits a distinguishable overall trend. We identified the closest fitting function for the MOS alone as

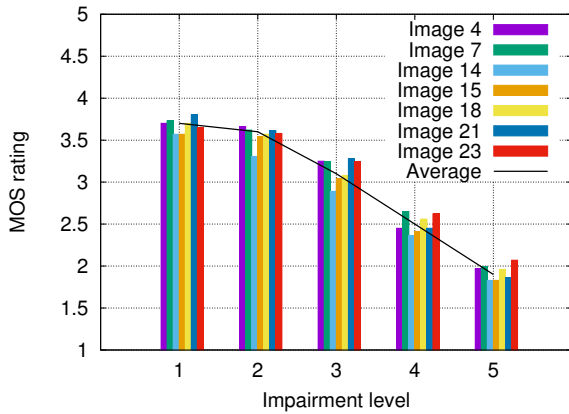
$$\bar{Q}_i^l = 0.4114 \cdot (\bar{M}_i^l)^{1.8217}. \quad (3)$$



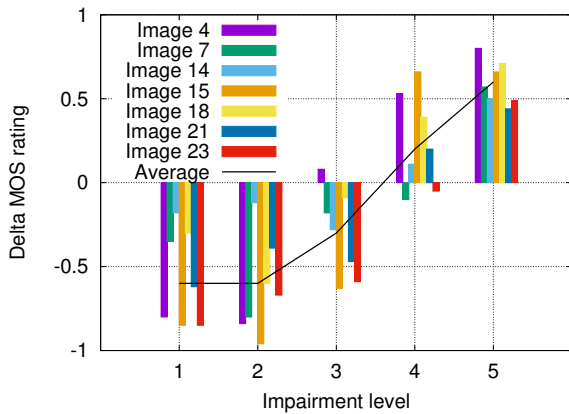
While limited to the image set evaluated herein, the visibly narrow distribution of the ratings illustrated in Fig. 5 indicates that other images might fall into this trend as well. In turn, predictions of MOS ratings for different presentation modes become likely, with the caveat that in the current setup, the real-world backdrop was fairly static.



(a) Mean opinion scores from the augmented vision presentation mode.

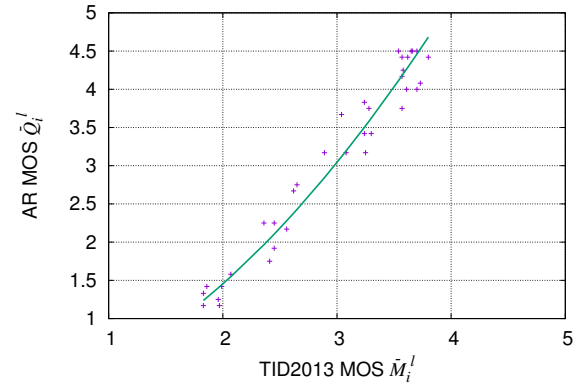


(b) Mean opinion scores as reported for the TID2013 data set.



(c) Difference in mean opinion scores for impairment levels between traditional opaque and augmented reality presentation modes.

**Fig. 4** Overview of impairment relationship to mean opinion scores for augmented vision and opaque presentation modes [28].



**Fig. 5** Scatter plot of mean opinion scores in augmented setting (AR-MOS) and traditional opaque presentation (TID-MOS), including functional relationship [28].

#### 4 Augmenting QoE and QoS

We initially evaluate the Pearson correlations for the AR-MOS  $\bar{Q}_i^l = f(\bar{M}_i^l)$ , the differential AR-MOS  $\bar{Q}_i^l = f(\bar{M}_i^l)$ , and the TID2013 MOS  $\bar{M}_i^l$  with respect to different image quality metrics. Specifically, we consider the Feature Similarity Index for color images [37], the Structural Similarity index (SSIM) and its scaled version (MSSIM) [38] next to the Peak Signal to Noise Ratio (PSNR). As a more detailed discussion of these metrics is beyond the scope of this article, we refer the interested reader to, e.g., [36], for an initial discussion of the impact of these metrics in a traditional setting. The results are provided in Table 1 for Pearson and Spearman correlations.

We note that almost all combinations of objective image quality metric values and MOS ratings exhibit a correlation coefficient above 0.8. Spearman and Pearson correlations are fairly comparable overall. The latter is commonly slightly higher in our evaluations, whereas it is slightly lower for the TID data set. We furthermore note that all correlations were significant at the  $p < 0.01$  level. We caution, however, that the image subset will have an impact on these values, especially when compared to the complete TID2013 data set: With the employment of the different image impairments on a larger image set, these values are likely to change.

**Table 1** Overview of Spearman and Pearson correlation coefficients between objective quality metrics (QoS) and determined opinion scores (QoE). All have a significance level of  $p < 0.01$ .

Type	Metric	AR-MOS $\bar{Q}_i^l$	Diff. AR-MOS $\bar{Q}_i^l$	TID2013 MOS $\bar{M}_i^l$
Pearson	FSIMc	0.937	0.951	0.947
	MSSIM	0.943	0.936	0.951
	PSNR	0.894	0.826	0.905
	SSIM	0.927	0.882	0.942
Spearman	FSIMc	0.914	0.916	0.958
	MSSIM	0.915	0.903	0.958
	PSNR	0.895	0.822	0.893
	SSIM	0.916	0.878	0.955

Overall, however, we note a similar trend to the one described in [36], where the PSNR does not result in the highest correlation between experimentally determined MOS values and objectively calculated fidelity metrics. We nevertheless note that a correlation of above 0.89 might still enable a useful first comparison if computationally undemanding metrics are required.

Next, we note a jump when considering structural similarity metrics, which have a significantly increased correlation with the MOS and are generally comparable in their levels. This slightly changes when considering the differential AR-MOS, where the MSSIM and FSIMc are exhibiting a significantly stronger correlation with the user ratings than the other two objective metrics. Overall, we conclude that there seems to be a strong impact of (scaled) similarity metrics on the augmented vision QoE ratings by users.

#### 4.1 QoE and QoS Relationships in Augmented Vision Scenarios

We now integrate the presented relationships of the objective image quality metrics as QoS indicators and the resulting MOS values as QoE indicators. Current approaches in QoE evaluations that describe this relationship include the exponential interdependency of QoE and QoS (IQX) and the Weber–Fechner Law (WFL) integrations, see, e.g., [16, 17] for a more in–depth coverage of the approaches. We initially set the QoS factors as derivatives from the objective image quality metrics as 1–FSIMc, 1–MSSIM, 1/PSNR, and 1–SSIM and insert them into the generalizing IQX formula of:

$$\text{QoE} = \alpha \cdot \exp(-\beta \cdot \text{QoS}) + 1. \quad (4)$$

This allows us to determine the factors  $\alpha$ ,  $\beta$  based on the experimentally determined MOS values for which we here employ the refined differential AR-MOS  $\bar{Q}'$ . Our reasoning to shift to the differential MOS is that, here, we are not focused on the comparison of the MOS with those reported for publicly available data sets, but we strive to determine the augmented vision scenarios' own underlying dynamics independently. We provide the factors for the different approaches, including the  $\chi^2$  sum of differences across the five impairment levels, in Table 2.

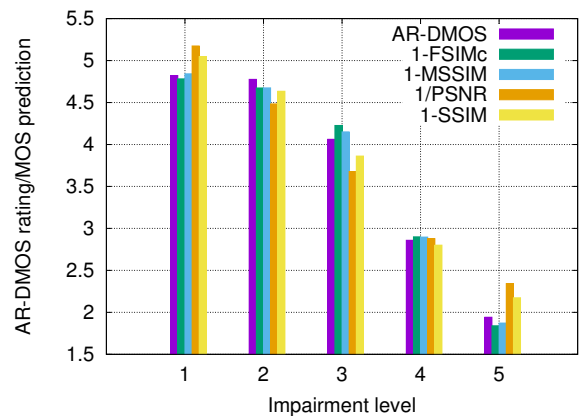
**Table 2** Solutions for parameters  $\alpha$ ,  $\beta$  for different QoS factors within the IQX approach and resulting  $\chi^2$  sum of errors.

QoS Factor	$\alpha$	$\beta$	$\chi^2$
1-FSIMc	3.869	7.674	0.051
1-MSSIM	4.032	11.400	0.025
1/PSNR	47.183	93.064	0.518
1-SSIM	4.891	4.599	0.168

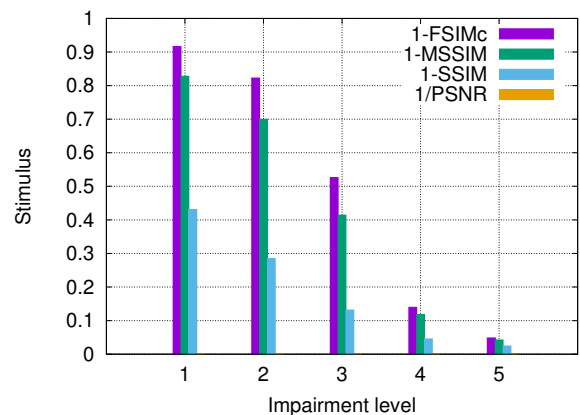
We observe that the smallest  $\chi^2$  value determined is for the MSSIM objective image quality metric. We illustrate the predicted QoE values and the experimentally determined MOS scores  $\bar{Q}'$  in Fig. 6. We initially observe that both FSIMc and MSSIM based QoS factors closely follow the MOS values' trend. Furthermore, we observe that the SSIM curve partially follows the trend of the MOS values, but to a lesser extent. Lastly, the inverse PSNR provides the visually highest deviation and presents a more linear approximation, which fails to capture the underlying behavior of  $\bar{Q}'$ . Next, we evaluate the WFL relationship for the determined parameters, which can be approximated (according to [17]) as

$$q(\text{QoS}) = \exp(\alpha \cdot \exp(-\beta \cdot \text{QoS}) - \alpha). \quad (5)$$

Thus, the stimulus-transforming CSTF function can be employed for the different metrics to determine their impacts. We illustrate the resulting WFL stimulus curves in Fig. 7. We immediately observe that in all cases, the stimulus function is monotonically declining with increasing impairments. We additionally note that the objective metric levels are ordered by their correlation with the QoE from experimentation, with notable declines in stimulus for the higher and lower ends of the impairment ranges evaluated.



**Fig. 6** Prediction of the QoE based on the IQX approach for different QoS metrics across all images [28].



**Fig. 7** Prediction of the stimulus according to the WFL approach employing different metrics [28].

This clearly indicates that operational points on either ends will suffer from diminishing returns of optimizations, which is a different relationship than observed in, e.g., [17].

## 5 Relationship of QoE and EEG

The overlay of content to be displayed with reality for a user of vision augmenting devices poses significant challenges for an evaluation of the impact of media quality. As the context of operation and background reality for a device change, the combined view that that a device operator's visual system faces cannot be subscribed ex-ante. In turn, the determination of the QoE that a user derives in such scenario is challenging. One potential alleviation to this problem could be the capturing of cognitive strain that emerges in these situations to approach the QoE, which we focus on in this section. We consider a subset of five subjects from the overall group of participants, who were outfitted with the wearable EEG head device described in Section 2. Four of the participants were male, with an overall average age of 29.4 years (SD=7.2 years) with two of the male subjects wearing prescription eye wear. As indicated in [29], the Low and High Gamma frequency ranges of the captured EEG data from the consumer-grade device indicate a promising relationship. We focus on the consideration of the Low Gamma range as it might relate to visual perception [39] and is reported from the device's drivers for the range from 31-39.75 Hz. As the device manufacturer's integrated chipset performs several operations on the original data before presenting values [40], "there is no longer a simple linear correlation to units of Volts." The reported values are, in turn, predominantly useful for qualitative analysis of relationships, which we focus on in the following.

We illustrate the behavior of  $\bar{E}^l(b,t)$  observed for the Low Gamma band  $b$  in Fig. 8a, whereby we compare the different image impairment levels  $l$  for the time that users were actively viewing the different images. We initially note that the lowest impairment level (i.e., highest QoE) initially results in the highest measured result. Other quality levels, however, are somewhat interspersed and out-of-order. For the majority of the impairment levels, we observe an initial change from the first seconds of viewing to the ordered result towards the end of the viewing period, whereby the last seconds of viewing exhibit a stable order between impairment levels. Interestingly, we notice a quality reversal effect, whereby the measured level of the Low Gamma signal is lower for the unimpaired original than for the lowest impaired level employed. We illustrate the Coefficient of Variation (CoV) based on the individual time and impairment level averages as metric for the variability of the obtained results in Fig. 8b. We notice that, in general, the level of variability is below one, with an approximate overall level around 0.6 to 0.7 or lower, which typically indicates an acceptable level of variability (especially considering our qualitative view on the results here).

Lastly, we evaluate the Pearson correlation to the impairment level subject ratings for the differential MOS in the AR

setting,  $\bar{Q}^l$ , with the obtained EEG signal averages, illustrated in Fig. 8c. We notice that the quality-level based correlation is slightly higher for short and long viewing period averages than the correlation to the differential MOS, but exhibits a significant "slump" in the medium viewing time periods. The linear dependency of the EEG averages for the Low Gamma band and the subject-reported QoE, on the other hand, remains well above 0.85 throughout the viewing period, indicating an underlying relationship between the QoE and EEG levels.

## 6 Discussion

We now discuss the highlights of our findings and their implications in greater detail. We denote the MOS difference described in Section 3 as the Visual User Experience Difference (VUED) – This effect indicates that the shift in presentation mode results in an initial increase in the QoE as indicated by the MOS in an AR setting (as well as its differential counterpart). As impairments increase, the differences in the reported QoE decrease. The MOS  $\bar{Q}_i^l$  exhibits a reversal effect, hereby the opaque presentation becomes advantageous for the lower quality levels. For the differential version as reference, we would not observe such a reversal, as the inclusion of the  $\bar{Q}_i^0$  ratings results in an always advantageous augmented vision presentation (i.e., it "shifts" the MOS). These results can be explained by the presentation that overlays with reality and obfuscates or diminishes some of the visual impairments that would be noticeable in an opaque setting. As the impairments increase in number and coarseness, they become visually more prominent in the sight overlay and result in less favorable ratings. In comparison with our earlier observations made for video in similar settings, see [41], we note that an emerging trend seems to indicate that little benefit can be attained with respect to an increase of the perceptually-oriented QoE for the low impairment levels. Indeed, it seems from our current data set comparisons that no significant gains can be achieved for increased image qualities at the high end. While this behavior is in line with the IQX and WFL theories observed earlier for opaque settings, the impacts we observe here are predominantly situated in the range of high, medium, and low quality ranges, whereby very high and very low impairments of the source material would both cause a lower change in stimulus and the resulting overall MOS in augmented vision scenarios. As the subjective QoE is even more complicated to derive when the background of displayed content dynamically changes (such as in augmented binocular vision scenarios), we employ contemporary EEG measurements to evaluate its usefulness in approximations of the QoE. We found that with significant viewing time for a smaller subject pool, the level of Low Gamma frequency bands is highly correlated to the actual image quality levels that subjects selected as well as the actual image quality level. This relationship opens significant potentials in future approximations for the QoE.

We find significant impacts for future media presentations employing vision augmenting mechanisms, such as those in common augmented reality applications:

- There is little benefit to increase the quality of images to be presented on vision augmenting devices beyond a “good” rating, as there is negligible impact on the QoE.
- The move from opaque presentation to a see-through presentation format results in a shift for different levels of impairments, an effect that we label VUED.
- Considering the differential MOS in the augmented scenarios results in a strong support for maintaining the current approach of employing objective image quality metrics as QoS factors to approximate the QoE, with similar preferences for individual metrics.
- Additional comparisons with current theories in QoE indicate that, while the findings for augmented vision settings are compatible to prior works, there is a shift in how the stimulus operates on the higher and lower ends.
- EEG measurements indicate that a direct relationship between media consumption duration and quality level exists, opening potential future approximation avenues.

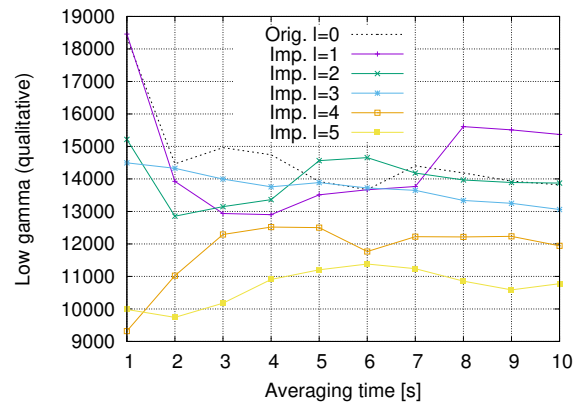
An initial concluding guideline based on our findings is to focus on the medium ranges of objective image qualities.

## 7 Conclusion

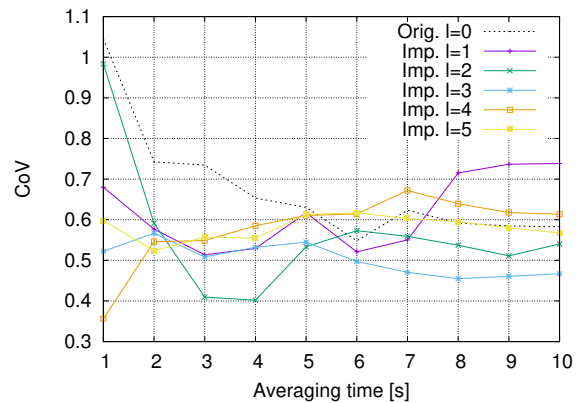
Despite augmented reality entering practical applications in professional and consumer spaces, little is known about the QoE in these scenarios. Throughout this article, we derived an overview of comparative characteristics for an experimental evaluation of the QoE, which we contrast with a large public data set that provides ground truth. Our comparisons include (i) image quality metrics as QoS factors, (ii) subjective ratings expressed by MOS as QoE indicators, and (iii) EEG scans of the Low Gamma frequency band as a new approach to the determination of the QoE. The configuration of optical see-through devices and their utilization scenarios make continuous user quality ratings a difficult task. With constantly changing backgrounds, rendering specific color(s) invisible or less discernible, employing EEG measurements as an approach that does not explicitly rely on user ratings, but enables a connection to the media quality appears as a promising avenue for future research. Additional finer-grained and absolute value measurements of the EEG are required for a broader variety of experiments to ultimately enable a direct relationship between EEG signals, QoE, and QoS, including multiple EEG measurement points. Based on the initial qualitative results that we provided in this article, we are encouraged to continue along this trajectory in our ongoing and future research efforts.

With the described emerging VUED relationships of QoE and QoS in vision augmenting settings, we envision that future

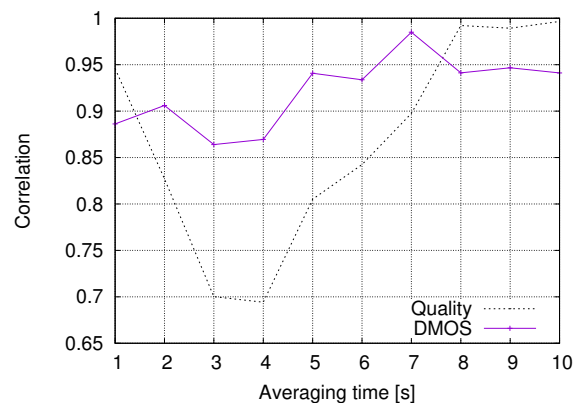
research will enable a more detailed view on the media qualities to provide the community with a ground truth data set. The implications of a more direct relationship of image quality ratings as MOS and QoE can be manifold, especially for future QoE predictions in augmented reality settings prior to sending different levels of impaired image data to mobile vision augmenting devices.



(a) Averages over different time spans from initial view of the media.



(b) Coefficient of Variation averages from initial view of the media.



(c) Correlations with image quality levels.

Fig. 8 Low Gamma EEG frequency band characteristics.

## Acknowledgement

This work was supported in part by a Faculty Research and Creative Endeavors grant from Central Michigan University.



## References

- [1] Van Krevelen, D.W.F., Poelman, R. "A Survey of Augmented Reality Technologies, Applications and Limitations." *The International Journal of Virtual Reality*. 9(2), pp. 1–20. 2010.
- [2] Kanbara, M., Okuma, T., Takemura, H. "A stereoscopic video see-through augmented reality system based on real-time vision-based registration." *Proceedings of the IEEE Virtual Reality Conference*. pp. 255–262. 2000.  
<https://doi.org/10.1109/VR.2000.840506>
- [3] Cisco, Inc., "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019." *Technical Report*. 2015.
- [4] E. Kruijff, E., Swan II, J. E., Feiner, S. "Perceptual issues in augmented reality revisited." *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. pp. 3–12. 2010.  
<https://doi.org/10.1109/ISMAR.2010.5643530>
- [5] Bonanni, L., Lee, C. H., Selker, T. "Attention-based design of augmented reality interfaces." *Proceedings of the ACM Human Factors in Computing Systems Conference (CHI)*. pp. 1228–1231. 2005.  
<https://doi.org/10.1145/1056808.1056883>
- [6] Kalkofen, D., Mendez, E., Schmalstieg, D. "Interactive focus and context visualization for augmented reality." *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. pp. 191–201. 2007.  
<https://doi.org/10.1109/ISMAR.2007.4538846>
- [7] Sanches, S. R. R., Tokunaga, D. M., Silva, V. F., Tori, R. "Subjective video quality assessment in segmentation for augmented reality applications." *Proceedings of the Symposium on Virtual and Augmented Reality (SVR)*. pp. 46–55. 2012.  
<https://doi.org/10.1109/SVR.2012.2>
- [8] Klein, E., Swan, J. E., Schmidt, G. S., Livingston, M. A., Staat, O. G. "Measurement Protocols for Medium-Field Distance Perception in Large Screen Immersive Displays." *Proceedings of the IEEE Virtual Reality Conference (VR)*. pp. 107–113. 2009.  
<https://doi.org/10.1109/VR.2009.4811007>
- [9] Livingston, M. A., Barrow, J. H., Sibley, C. M. "Quantification of Contrast Sensitivity and Color Perception using Head-worn Augmented Reality Displays." *Proceedings of the IEEE Virtual Reality Conference (VR)*. pp. 115–122. 2009.  
<https://doi.org/10.1109/VR.2009.4811009>
- [10] Woods, R. L., Fetchenheuer, I., Vargas-Martín, F. and Peli, E. "The impact of non-immersive head-mounted displays (HMDs) on the visual field." *Journal of the Society for Information Display*. 11(1), pp. 191–198. 2003.  
<https://doi.org/10.1889/1.1831704>
- [11] Seeling, P., Reisslein, M. "Video traffic characteristics of modern encoding standards: H. 264/avc with svc and mvc extensions and h. 265/hevc." *The Scientific World Journal*. 2014(189481), pp. 1–16. 2014.  
<https://doi.org/10.1155/2014/189481>
- [12] Le Callet, P., Möller, S., Perkiš, A. (Eds.). "Qualinet white paper on definitions of quality of experience." *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*. pp. 1–24. 2013.
- [13] Calvo, R. A., D’Mello, S. "Affect detection: An interdisciplinary review of models, methods, and their applications." *IEEE Transactions on Affective Computing*. 1(1), pp. 18–37. 2010.  
<https://doi.org/10.1109/T-AFFC.2010.1>
- [14] Hektner, J. M., Schmidt, J. A., Csikszentmihalyi, M. "Experience sampling method: Measuring the quality of everyday life". *Sage*. 2007.
- [15] ITU-R, 1534-1. "Method for the subjective assessment of intermediate quality levels of coding systems (MUSHRA)." *International Telecommunication Union*. 2003.
- [16] Fiedler, M., Hossfeld, T., Tran-Gia, P. "A generic quantitative relationship between quality of experience and quality of service". *IEEE Network*. 24(2), pp. 36–41. 2010.  
<https://doi.org/10.1109/MNET.2010.5430142>
- [17] Reichl, P., B. Tuffin, B., Schatz, R. "Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience." *Telecommunication Systems*. 52(2), pp. 587–600. 2013.  
<https://doi.org/10.1007/s11235-011-9503-7>
- [18] Gabbard, J. L., Fitch, G. M., Kim, H. "Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications." *Proceedings of the IEEE*. 102(2), pp. 124–136. 2014.  
<https://doi.org/10.1109/JPROC.2013.2294642>
- [19] Scholler, S., Bosse, S., Treder, M. S., Blankertz, B., Curio, G., Mueller, K.-R., Wiegand, T. "Toward a direct measure of video quality perception using eeg." *IEEE Transactions on Image Processing*. 21(5), pp. 2619–2629. 2012.  
<https://doi.org/10.1109/TIP.2012.2187672>
- [20] Sweller, J. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science*. 12(2), pp. 257–285. 1988.  
[https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- [21] Gevins, A., Smith, M. E., McEvoy, L., Yu, D. "High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice." *Cerebral Cortex*. 7(4), pp. 374–385. 1997.  
<https://doi.org/10.1093/cercor/7.4.374>
- [22] Backs, R. W. and Boucsein, W. "Engineering psychophysiology: issues and applications." *Lawrence Erlbaum*. 2000.
- [23] Jang, Y., Woo, W. "Unified Visual Perception Model for context-aware wearable AR." *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. pp. 1–4. 2013.  
<https://doi.org/10.1109/ISMAR.2013.6671818>
- [24] Bertuccelli, L. F., Khawaja, T., Walker, B. N., O’Neill, P. "Ongoing development of a user-centered, ar testbed in industry." *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. pp. 253–254. 2014.  
<https://doi.org/10.1109/ISMAR.2014.6948436>
- [25] Mueller, M. M., Gruber, T., Keil, A. "Modulation of induced gamma band activity in the human eeg by attention and visual information processing." *International Journal of Psychophysiology*. 38(3), pp. 283 – 299. 2000.
- [26] Demiralp, T., Bayraktaroglu, Z., Lenz, D., Junge, S., Busch, N. A., Maess, B., Ergen, M., Herrmann, C. S. "Gamma amplitudes are coupled to theta phase in human eeg during visual perception." *International Journal of Psychophysiology*. 64(1), pp. 24 – 30. 2007.  
<https://doi.org/10.1016/j.ijpsycho.2006.07.005>
- [27] McFarland, D. J., Parvaz, M. A., Sarnacki, W. A., Goldstein, R. Z., Wolpaw, J. R. "Prediction of subjective ratings of emotional pictures by eeg features." *Journal of Neural Engineering*. 14(1), p. 016009. 2017.  
<https://doi.org/10.1088/1741-2552/14/1/016009>
- [28] Seeling, P. "Visual user experience difference: Image compression impacts on the quality of experience in augmented binocular vision." *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC)*. 2016.  
<https://doi.org/10.1109/CCNC.2016.7444911>
- [29] Seeling, P. "Augmented vision and quality of experience assessment: Towards a unified evaluation framework." *Proceedings of the IEEE ICC Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*. 2015.  
<https://doi.org/10.1109/ICCW.2015.7247431>

- [30] ITU-T, "Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications." *International Telecommunication Union*. 2008.
- [31] ITU-R, "Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures." *International Telecommunication Union*. 2012.
- [32] Erikson, E. H. "*Identity: Youth and crisis*." WW Norton & Company. 1968.
- [33] Shirazi, A. S., Hassib, M., Henze, N., Schmidt, A., Kunze, K. "What's on your mind?: Mental task awareness using single electrode brain computer interfaces." *Proceedings of the ACM Augmented Human Conference*. pp. 1–4. 2014.  
<https://doi.org/10.1145/2582051.2582096>
- [34] Mak, J. N., Chan, R. H. M., Wong, S. W. H. "Evaluation of mental workload in visual–motor task: Spectral analysis of single-channel frontal eeg." *Proceedings of the Annual Conference of the IEEE Industrial Electronics Society (IECON)*. pp. 8426–8430. 2013.  
<https://doi.org/10.1109/IECON.2013.6700546>
- [35] Winkler, S. "Analysis of Public Image and Video Databases for Quality Assessment." *IEEE Journal of Selected Topics in Signal Processing (JSAC)*. 6(6), pp. 616–625. 2012.  
<https://doi.org/10.1109/JSTSP.2012.2215007>
- [36] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., Kuo, C.-C. J. "Image database TID2013: Peculiarities, results and perspectives." *Signal Processing: Image Communication*. 30, pp. 57–77. 2015.  
<https://doi.org/10.1016/j.image.2014.10.009>
- [37] Zhang, L., Zhang, L., Mou, X., Zhang, D. "FSIM: A Feature Similarity Index for Image Quality Assessment." *IEEE Transactions on Image Processing*. 20(8), pp. 2378–2386. 2011.  
<https://doi.org/10.1109/TIP.2011.2109730>
- [38] Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E. "Image quality assessment: from error visibility to structural similarity." *IEEE Transactions on Image Processing*, 13(4), pp. 600–612. 2004.  
<https://doi.org/10.1109/TIP.2003.819861>
- [39] Buzsaki, G. "*Rhythms of the Brain*." Oxford University Press. 2006.
- [40] Eeg band power values: Units, amplitudes, and meaning. [Online]. Available from: <http://support.neurosky.com/kb/development-2/eeg-band-power-values-units-amplitudes-and-meaning> [Accessed: 31st August 2017]
- [41] Seeling, P. "Towards quality of experience determination for video in augmented binocular vision scenarios." *Signal Processing: Image Communication*. 33(0), pp. 41–50. 2015.  
<https://doi.org/10.1016/j.image.2015.02.006>