Subjective Study of 3D Mesh Quality Scores in Virtual Reality

Dalia A. ALfarasani
Yu-Kun Lai
Paul L. Rosin
School of Computer Science & Informatics, Cardiff University
UK

ABSTRACT

3D meshes are widely used in Virtual Reality as essential graphical elements for creating immersive virtual environments. In practice, the 3D meshes being used are often subject to some manipulations, where some details may be lost and some noise could be introduced, e.g., due to the limited transmission bandwidth. While existing studies have considered 3D mesh quality measures in the desktop setting, we consider how different 3D distortion types affect the perceptual quality of 3D shapes when viewed in a Virtual Reality setup (with users wearing a Meta/Oculus headset). Our experiment collected mean opinion scores (MOS) for each distorted shape by showing both distorted and reference shapes to users, and compared the results with meshes viewed in the traditional display. This paper aims to understand the effect of different types and levels of 3D mesh distortions on perceived quality and user experience in VR. We analyse correlations of two settings (VR and desktop) using the Pearson and Spearman correlation coefficients, which show a positive relationship between the two settings. However, in virtual reality, perception appears more sensitive to particular distortions than others, compared with the desktop setting, which can provide helpful guidance for downstream applications.

CCS CONCEPTS

• **Computing methodologies → Virtual reality**.

KEYWORDS

Subjective study, Quality assessment, 3D mesh quality, Distortions type

ACM Reference Format:
Dalia A. ALfarasani, Yu-Kun Lai, and Paul L. Rosin. 2023. Subjective Study of 3D Mesh Quality Scores in Virtual Reality. In *2023 the 7th International Conference on Virtual and Augmented Reality Simulations (ICVARS) (ICVARS 2023), March 03–05, 2023, Sydney, Australia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3603421.3603423

1 INTRODUCTION

Recent advances in 3D mesh modelling, representation, and rendering have progressed to the point that they are now extensively employed in many applications, such as networked 3D gaming, 3D virtual reality (VR) and augmented reality (AR). With VR, users can experience high-quality, immersive virtual environments in real-time using the latest advancements in computer graphics hardware and software [29]. Increasing the visual quality of a mesh by using a large number of vertices and faces provides a more detailed representation. However, the added complexity leads to increased requirements for data storage, processing power (CPU and GPU), and network bandwidth, especially for real-time applications and when data needs to be transmitted over the network. As a result, a trade-off between graphical model visual quality and processing time frequently arises, necessitating determining the quality of 3D graphic resources.

Several geometric modifications may be applied to 3D mesh models like compression, simplification, and watermarking. These processing procedures may influence the appearance and visual quality of the 3D models and, as a result, the quality of the user experience (QoE). Thus, subjective quality evaluation tests are essential for evaluating visual quality as perceived by human observers. Subjective methods involve a group of human participants being asked to rate the quality of a collection of 3D meshes that have been subjected to different types and levels of distortion. The output of the subjective method is a set of Mean Opinion Scores (MOS), which enables predictive models to be developed and evaluated, taking subjective scores as ground truth. Some subjective methods can be used in 2D image and 3D graphical areas, for example, single stimulus, double stimulus, subjective assessment methodology for video quality, and pairwise comparison (PC) [2, 5, 22]. Nevertheless, choosing the appropriate subjective technique is not easy since we must verify that such methods produce accurate and reliable findings. Previous subjective tests in the field of computer graphics were conducted to evaluate the visual quality of static and animated 3D models [7, 9, 15]. As shown in most papers, there is no agreement on the appropriate approach to assessing the quality of 3D models [18]. As we focus on the human visual system (HVS), which is strongly linked with perceptual quality measures, we concentrate on perceived 3D mesh perceptual quality measures using a VR headset, not on purely geometric measurements that ignore human perception. Bulbul et al. [3], Lavoué and Mantiuk [15] and Muzahid et al. [17], are mostly working in the perceptual area, and provide reviews of more broad 3D visual quality evaluation techniques.

In our case, we used a pairwise comparison (PC) approach of 3D meshes, where participants were asked to rate a collection of different levels of 3D mesh distortion in terms of visual quality, compared with the reference undistorted shape, which is presented to the user along with the distorted shape. PC is simpler and more intuitive for users, which ensures that users can concentrate on judging the quality of the distorted shape in comparison to the given

ICVARS 2023, March 03–05, 2023, Sydney, Australia

ACM ISBN 978-1-4503-9746-9/23/03…$15.00
https://doi.org/10.1145/3603421.3603423

reference shape. We propose to measure how different distortions (noise, smoothing, etc.) of 3D shapes affect the perceptual quality of 3D objects in a VR environment by collecting subjective scores for distorted shapes. We compare the MOS between VR and desktop settings. Moreover, we analyse different 3D mesh distortion types with different 3D shapes to determine which distortion type/shape shows significant results. To make the comparison between VR and desktop settings easier, we used an existing database evaluated on the desktop display. As VR is becoming a popular way of consuming and visualising 3D content with high resolutions, we build an application to carry out VR experiments using Meta/Oculus Quest 2 as one of the most popular headsets.

2 RELATED WORK

Immersive virtual reality (IVR) enables users to establish deeper connections with 3D environments by integrating immersion as a significant element [16]. According to some scholars, IVR can be classified into low and high IVR. The low IVR only provides low immersion using techniques such as 2D computer screens, while the high IVR achieves high immersion, benefiting from techniques such as Head Mounted Displays (HMD) and Cave Automatic Virtual Environment (CAVE) [10]. Many subjectively rated databases have been released during the last few years. Winkler et al. [28] provided an overview and a comparison of publicly available image databases. The first studies on subjective quality assessment of 3D static meshes were led by Watson et al. [27], who tried to measure the visual fidelity of simplified meshes. Corsini et al. [7] attempted to adopt existing experimental protocols to the subjective quality assessment of watermarked 3D meshes. Lavoué et al. [13] used multi-linear regression to optimise the weights of numerous mesh descriptors, including curvature values, dihedral angles, and the geometric Laplacian. Chetouani et al. [6] suggested employing a Support Vector Regression (SVR) model to combine various commonly used full-reference quality measures to increase the correlation between prediction and human observations.

Torkhani et al. [23] utilised their proposed TPDM measure (Tensor-based Perceptual Distance Measure) and proposed a machine learning method to train an SVR model to determine the link between the distortion distribution and quality scores. The existing studies like Lavoué et al.'s MSDM2 [13], Wang et al.'s FMPD [25], and Váša and Rus' DAME [24] are strong predictors of visual quality. Váša and Rus [24] studied dihedral angle discrepancy, whereas Lavoué et al. [12] suggested metrics based on local variances in curvature statistics. Local changes of attribute values at the vertex or edge level are included in these metrics, which are subsequently aggregated into a global score. On the other hand, Corsini et al. [8], calculated global roughness values per model before working out global roughness differences as measures. Torkhani et al. [23] incorporated perceptually motivated methods such as visual masking, which are similar to bottom-up image quality measurements. There is existing research that utilised 3D models to estimate quality in 3D distorted meshes like [7, 12, 14, 22, 24], which conducted subjective assessments using 3D static or dynamic models, but none of them used a VR setting.

Bulbul et al. [3] conducted a subjective assessment survey to determine how downsampling or introducing coordinate noise to a 3D point cloud impacts perceived quality. They also present a subjective analysis of 3D point cloud denoising algorithms using the Double Stimulus-Impairment-Scale (DSIS) approach and the correlation with objective measures. However, this research aims to evaluate the performance of denoising methods rather than the quality of other types of distortions, such as compressed mesh geometry. Recently Nehme et al. [18] used a VR experiment to see how the explicit reference affects the quality evaluation of coloured 3D models. They conducted a psycho-visual study to compare the performance of two methods: ACR-HR (Absolute Category Rating with hidden references) and DSIS (with explicit references). They used two sets of observers, and two tests were given to each group in a different order. The experiment utilised the VR HTC Vive Pro in fixed position mode in an immersive virtual world. Their focus is to analyse the subjective quality assessment methods for coloured meshes, whereas our work measures subjective quality assessment for meshes with geometry only and compares the results with desktop setting.

3 SUBJECTIVE EXPERIMENT

In our experiment, we use a public database to evaluate the 3D mesh quality level of distortion using a VR headset (HMD). Figure 1 presents the LRIS/EPEL general-purpose database, which contains four reference models (Armadillo, Venus, Dyno and Rocker Arm) and 84 distorted meshes (21 distorted meshes for each reference mesh) [14]. This dataset used two different types of degradation, noise and Taubin smoothing [21], to simulate typical degradation of mesh quality due to e.g., compression and watermarking [14]. These distortions have different levels of strength and four types of locations on meshes: uniformly on the whole mesh, smooth areas, rough areas, intermediate areas, where different areas are identified based on local curvature variations. Note that Taubin smoothing is not applied to the smooth areas as the effect is hard to notice. For noise addition, subjective quality scores are provided for each distorted mesh in the form of MOS, ranging from 0 (worst quality) to 10 (best quality). Lavoué et al. [14] created noise by altering the three coordinates of the mesh's vertices with a randomly calculated offset between 0 and the specified maximum deviation. Smoothing was accomplished by applying Taubin [21] smoothing filter to the mesh's vertices.

These distortions were applied at three distinct intensities (visually selected): high, medium, and low (these levels correlate to the number of smoothing iterations and the maximum deviation value for noise addition). Finally, these distortions were applied in different locations on the meshes: evenly (across the whole object), only to smooth areas, rough regions, and intermediate regions. Each model generated 21 degraded versions: three noise strengths in four types of locations and three Taubin smoothing strengths in three types of locations (i.e., excluding smooth areas).

In this work, we compare VR and desktop settings. In the desktop setting experiment, Lavoué et al. [14] used a 1280×720 resolution monitor, such that different types and levels of distortions could be observed, and each model was displayed in a 600×600-pixel window. All models have a resolution of between 50,000 and 100,000 triangles, so the details of the model can be viewed well. They used rotation, interaction and zoom operations to allow the participant

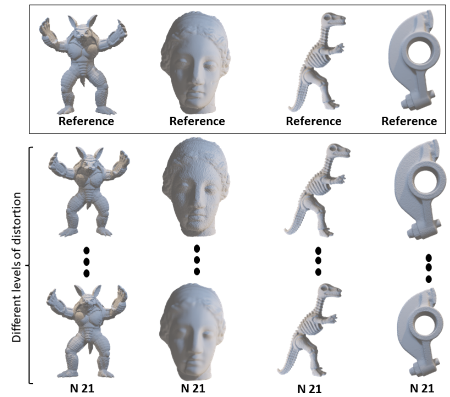


Figure 1: Examples of 3D meshes belonging to the LIRIS/EPFL General-Purpose database. The top row shows the four reference meshes, and the remaining rows show that we have 21 distorted models for each shape, so the total number of shapes is 88.

to interact with the model (e.g. mouse clicks) in their experiment. Also, [14] showed the models (both reference and distorted versions) to the participants on the desktop setting. The participants were allowed to browse through shapes so that they could memorise the worst/best quality shapes.

Our study uses a VR setting that does not allow participants to see all the models simultaneously. However, we show participants a trailer with a different dataset to make sure the participant has an idea of how the experiment will be. These models were obtained from different sources which used different scanners. For example, the Armadillo model is a manifold/simplified version of the original model that was created from scanning data by the Stanford Computer Graphics Laboratory. The Dinosaur model is courtesy of Cyberware Inc. The Venus and Rocker Arm models are courtesy of the AIM@SHAPE project. Our subjective study was conducted using a pairwise comparison (PC) method in a VR setting.

3.1 Evaluation methodology

In our experiment, we used a PC approach to evaluate the quality of distorted meshes with respect to the reference (undistorted) 3D mesh. More precisely, throughout the experiment, each participant was provided with a distorted version to compare against the reference version. Then, each participant was asked to measure the quality of the distorted version compared with the reference by using a slider bar (0 worst quality, 10 best quality). Such PC is simpler to perform and requires less mental effort from the participants than DSIS, ACR-HR, etc. This experiment shows a new approach to using a VR setting to compare how different platforms (VR versus a desktop display) affect perceived mesh quality, which has not been done before. We will provide a new comparison approach between VR and desktop settings to simulate real environments and identify similarities and dissimilarities between human perception in these settings.

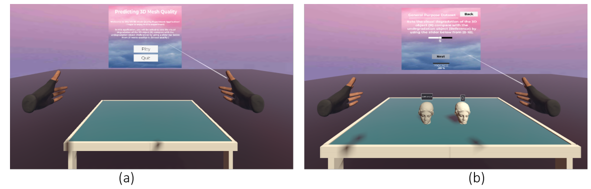


Figure 2: (a) Example of the home page showing our experimental environment and (b) example of the main page. The user can interact with the models in the experiment environment. The reference model is on the left, and the distorted model (B) is on the right side.

3.2 Display

We have considered different types of VR devices such as Valve Index and HTC Vive, and Meta/Oculus Quest 2 has the highest market share with a decent resolution. This ensures the study is more consistent with the common user experience. In our experiment, the display technology consists of a Meta/Oculus Quest 2 HMD with Qualcomm Snapdragon XR2 Platform, a single Fast-Switch LCD 1832×1920 pixels per eye with refresh rate 72Hz and tracking inside and outside 6 DOF (degrees of freedom). The experiment was built as an application in Unity3D and rendered with a resolution of 1832×1920, as shown in Figure 2. The experiment is based on the PC method; the participants rate the quality between two models where one is the (undistorted) reference and the other is a distorted version. Participants were allowed to explore the 3D mesh object by using touch controllers. In the experiment, the participant's head's position and rotation are used in the Unity3D application to provide a first-person perspective to explore the object [20]. Since depth perception could also play a significant role in the selection, we presented the objects equidistantly at (20cm) distance from the participants' eyes and they could freely move closer and further from the objects to explore them.

3.3 Participants and training

Before we started the experiment, we began with a trial session, as recommended by the ITU-R BT.500 [19], in order that participants became acquainted with the virtual environment and task, such that they fully comprehend the experiment's task. This stimulus outcome was not recorded. The main reason for this trailer is to enable participants who are not familiar with VR devices to learn how to rotate, scale and transform 3D objects. The experiment was conducted at Cardiff University and involved students aged between 20 and 40, with twenty females and thirty males. All participants reported normal or corrected-to-normal vision.

3.4 Procedure

The experiment used a Meta/Oculus Quest 2 headset, and asked participants to rate the quality of the 3D distorted models. The experiment showed several 3D scanned meshes from the LIRIS General-purpose database, along with distorted versions. Every participant was expected to take 20-25 minutes to finish viewing the models. The experiment was carried out through a computer application that presents the 3D objects on the VR headset in a random order, with paired 3D objects appearing side by side. At each comparison, participants compare the reference model with the distorted model by using a VR touch controller to scroll the slider bar score as illustrated in Figure 2. This way allows us to collect the MOS scores and analyse correlations of individual distortion types by using Pearson and Spearman coefficient correlation. The settings used in an experiment are critical because they can bias the results significantly, especially for computer-generated stimuli, where almost every element can be controlled.

3.5 Duration

The overall length of the experiment affects the efficiency of the experimental method, especially in VR where most of the subjects have not used the VR headset before and tend to exhibit symptoms of cybersickness both during and after the virtual environment experience [11]. To avoid these issues, we chose to display the reference and the test stimulus simultaneously side by side in the same scene. In this way, the number of presentations is halved. To avoid fatigue, boredom and cybersickness, we allow the participant to move around the lab or sit in a chair to reduce any motion sickness. Each subject's session took place on a single day in order to prevent any learning effect between stimuli. The stimuli were displayed in a random order (i.e. reference models, distortion types and distortion levels) to each participant. Each stimulus was presented once; the participant was not able to replay the scene.

4 DATA ANALYSIS

4.1 Screening participants and computing mean opinion scores (MOS)

Before performing any data analysis, we have tested the participants' performance to ensure the collected data is meaningful. We follow the ITU-R BT.500-13 recommendation [19], where we show a trailer with a different dataset to the participants to make sure they understand how the experiment works. Computing the Interquartile Range (IQR) [19] of our data, we identified outliers as values either greater than the third quartile plus 1.5×IQR or less than the first quartile minus 1.5×IQR. One outlying participant was found in both settings, and was rejected from the dataset.

To analyse user ratings, a common method is to compute the MOS for each stimulus.

$$MOS_e = \frac{1}{10 \times N} \sum_{i=1}^{N} s_{ie}, \tag{1}$$

where s_{ie} refers to the score assigned by participant i to the stimulus e, and N denotes the number of (valid) subjects. We further divided the scores by 10 to normalize them in the range of $[0, 1]$. We follow most of the existing work [1, 4] and set the scores such that 0 means the worst quality, and 1 is the best quality. So we expect the MOS to decrease as the distortion level increases. In Armadillo, we notice a strong consistency between the VR and desktop settings, as the participants in both settings showed almost the same behaviour for each type of distortion. However, for the rest of the models, we may see some disparities in the rating scores of the two settings. In fact, in some cases, desktop participants' scores are not consistent with the stimuli, i.e., the quality does not always drop when the level of distortion increases (e.g. for the Venus model), but VR viewers give accurate ratings.

Furthermore, we found that VR observers were able to detect several distortions that desktop observers missed. These initial findings suggest some discrepancies in the human perception w.r.t. different display techniques. In the next section, we will examine whether these differences are statistically significant and seek to explain their origins.

4.2 Observer agreement analysis

Prior to analysing the experiment's outcomes, it is critical to assess the subjects' agreement and check if they remained attentive during the experiment. We calculated the standard deviation σ values of the MOS scores from both the VR and desktop settings. In the VR setting $\sigma_{VR} = 0.0220$ whereas in the desktop setting $\sigma_D = 0.0278$. Since the standard deviation in the VR setting σ_{VR} is lower than that of the desktop setting σ_D, it shows that the user ratings are more consistent for the VR setting.

In order to further analyse the similarity and dissimilarity between subjective mesh quality, we look at the correlations between subject evaluations. To begin with, we examine the correlation of the VR setting and desktop setting for distortions of each 3D object of the dataset. We also check the correlation for each type of distortion. In this paper, the following two measures are used to measure the correlation between VR and desktop settings. The Pearson linear correlation coefficient (PLCC or r_p) measures the prediction accuracy of MOS, while the Spearman rank-order correlation coefficient (SROCC or r_s) measures the prediction monotonicity [26]. Both values of PLCC and SROCC range from -1 to 1, where 1 indicates a total positive correlation, -1 indicates a total negative correlation, and 0 indicates no correlation. Suppose in our case we have two (VR & desktop) settings $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$, both containing n values. The Pearson linear correlation coefficient r_p between settings x and y is calculated as follows.

$$r_p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the mean values of x and y. MOS scores x and y are sorted in the same order (either ascending or descending). Let X_i be the rank of x_i in x, and Y_i be the rank of y_i in y. We generate two new sequences $X = \{X_1, X_2, \ldots, X_n\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$.

Let $d_i = X_i - Y_i$, the Spearman rank-order correlation coefficient r_s between settings x and y is calculated as follows.

$$r_s = 1 - \frac{6\sum_{i=1}^{n}{}^{2}}{d}{}_{i}\, n(n^2 - 1) \tag{4}$$

5 RESULTS

In this section, we divide our results into two parts, distortion by shapes as illustrated in Figure 3 and distortion by types as illustrated in Figure 4.

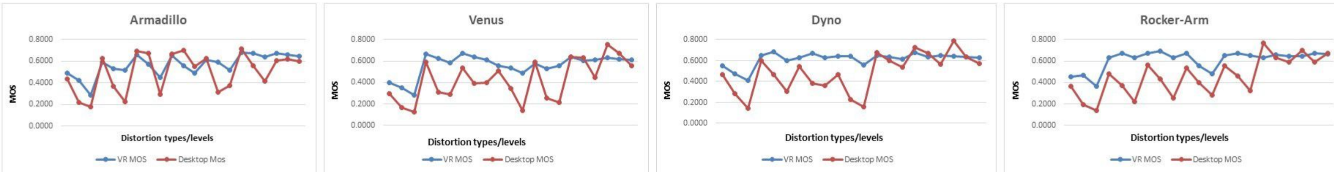


Figure 3: Comparison of MOS for both VR and desktop settings in the pairwise comparison (PC) experiment for all the stimuli shapes (the blue curve is for VR and red for desktop).

5.1 Distortion by shape

We now analyse MOS scores between VR and desktop settings on the basis of individual test shapes. Since the two modes of display have their own characteristics, as illustrated in Figure 3, the perceived quality is more consistent on some shapes than others. The x-axis corresponds to different distortion types, locations and strengths, and the y-axis shows the (normalised) MOS scores. We can see that Armadillo has the highest correlations in both Pearson linear coefficient correlation (PLCC) and Spearman coefficient correlation (SROCC) ($r_p = 0.754$, $r_s = 0.685$) compared with the rest of the models.

Armadillo contains some details, which means it is easy for participants to notice the different quality between the reference model and the distorted version. In comparison, the Venus model has a lower linear relationship and a higher nonlinear relationship ($r_p = 0.698$, $r_s = 0.701$) because a participant in the VR setting is easy to zoom and rotate the model compared with the desktop, which helps the participant to notice small areas that might not appear well on the desktop. The third model is Rocker Arm, which shows a fair relationship between VR setting and desktop setting ($r_p = 0.623$, $r_s = 0.513$). In both settings, participants often did not notice if there was a distortion in the shape because the shape did not have much details they could detect. The last model is Dyno which has the worst result compared with the rest of the models. The linear coefficient correlation is better than nonlinear coefficient correlation ($r_p = 0.624$, $r_s = 0.499$). The reason behind the worst correlation is that the Dyno model does not have a smooth area which makes it hard to detect the quality even if the reference is available. All the correlation results are listed in Table 1.

5.2 Distortion by type and location

As we explained in the previous section, this dataset has two main distortion types: adding noise and Taubin (smoothing), but each type has different levels of strength and different locations. We now group the results based on type and location (see Figure 4), and for each type, show different distorted shapes and levels of distortion strength. The PLCC and SROCC correlations between VR and desktop settings for different distortion types and locations are summarized in Table 2, with a detailed breakdown given in Table 1. People in the VR settings perceived differently compared to the desktop setting. As the VR has a higher resolution which makes it easy to interact with models, as a result, the MOS scores appear to be more sensitive to where distortions are applied. Firstly, Noise Uniform distortion is the most obvious distortion for the participants as noise is added uniformly to the whole shapes; they can easily notice and detect the distortion and the results are more consistent in both settings, as evidenced by high correlations ($r_p = 0.732$, $r_s = 0.753$).

For Noise Rough and Noise Intermediate, noise is only added to rough and intermediate regions. As these regions contain details already, added noise can be less noticeable. As shown in Figure 4, each group of 3 samples in the x-axis corresponds to three levels of distortion strength (low, medium and high) for each of the 4 shapes. Although adding more noise tends to lead to lower MOS scores in the VR settings, the drop is significantly smaller compared to adding noise uniformly. In the case of adding noise to the smooth regions, the MOS scores have larger drops with increasing strength of noise, close to the Noise Uniform case. This shows that with better observation/interaction, subjective scores are more sensitive to where distortion, especially noise, is applied. In contrast, the results of the desktop setting show little difference between locations.

A visual example is shown in Figure 5 where (b) is the shape with high-level noise applied on the rough regions, whereas (c) is with medium-level noise applied on the smooth regions. It is obvious that the distortion in (c) is more visible than (b), which is correctly reflected in the MOS scores in the VR setting, but not so in the desktop setting, where the strength of distortion rather than the location has more impact on the perceptual quality. Because of such differences, the correlations in these locations are significantly lower than in the uniform case.

We now compare smoothing (Taubin) and adding noise. As shown in Figure 4, MOS scores where smoothing is applied tend to be higher than with noise added, especially when distortions are applied uniformly. In contrast, in the desktop setting, these types of distortions have a similar impact. Similarly, different strength levels also have less impact on MOS scores, compared with the desktop setting. These also lead to lower correlations between VR and desktop settings in the case of smoothing. Nevertheless, if we consider all samples (shapes, distortions, locations and strength levels), the MOS scores remain highly correlated between VR and desktop settings ($r_p = 0.929$, $r_s = 0.929$).

Table 1: Pearson and Spearman correlation analysis comparing VR and desktop MOS scores for different stimuli (the distortion type followed by distortion location

Distortion Type and Location	Armadillo		Venus		Dyno		Rocker Arm	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
Noise Uniform	0.850	1	0.934	1	1	1	0.549	0.500
Noise Rough	0.993	1	0.882	1	0.657	0.500	0.093	0
Noise Intermediate	0.925	1	0.828	0.500	-0.348	0.500	0.729	0.500
Noise Smooth	0.667	0.500	0.981	1	0.667	0.866	0.995	1
Taubin Uniform	0.564	0.500	0.686	0.500	0.975	1	0.183	0.500
Taubin Rough	0.873	1	0.269	0.500	0.500	0.500	-0.660	-0.500
Taubin Intermediate	0.538	0.500	0.995	1	0.829	1	-0.868	-1
All	0.754	0.685	0.698	0.701	0.624	0.499	0.623	0.513

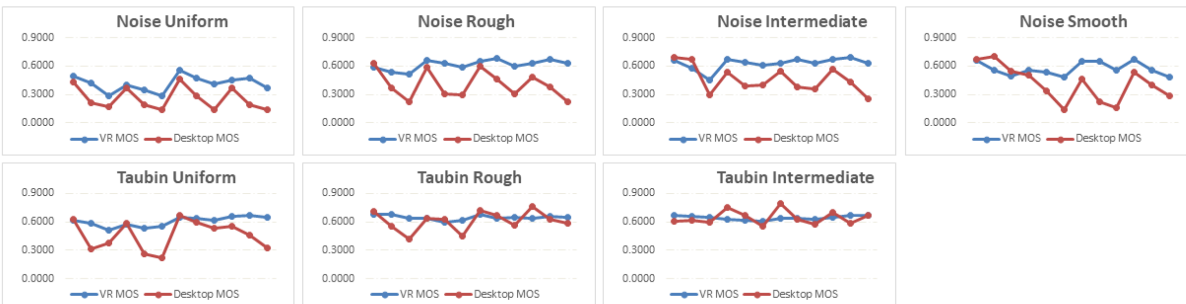


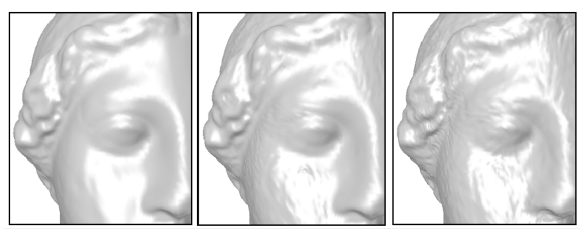
Figure 4: Comparison of MOS scores between the VR and desktop settings. Each subfigure shows a type of distortion (Noise or Taubin smoothing) applied to certain locations (Uniform, Rough, Intermediate and Smooth). The x-axis corresponds to each distorted shape (12 altogether for each subfigure corresponding to a combination of 4 shapes and 3 levels of distortion strength), and the y-axis shows the (normalized) MOS scores averaged over all subjects for the distorted shape. The blue and red curves correspond to the VR and desktop settings, respectively.

Table 2: Pearson and Spearman coefficient correlations between MOS scores from VR and desktop settings, grouped based on distortion types and locations.

Distortion Type/Location	Pearson Correlation (PLCC)	Spearman Correlation (SROCC)
Noise Uniform	0.732	0.753
Noise Rough	0.441	0.448
Noise Intermediate	0.273	0.266
Noise Smooth	0.335	0.387
Taubin Uniform	0.526	0.413
Taubin Rough	0.307	0.123
Taubin Intermediate	-0.078	-0.004
All	**0.929**	**0.929**

6 CONCLUSION

This paper proposed a subjective study using linear and nonlinear correlation coefficients to compare VR and desktop settings. Our analysis indicates the actual perceived mesh quality varies and is sensitive to different types/locations of distortion. We notice that overall MOS score distributions are highly correlated between the VR and desktop settings. However, in the VR setting, noise is more noticeable than a loss of details, when compared with the desktop setting. In particular, noise added to the entire shapes or smooth regions tends to be more noticeable than in other regions, and the differences are much more significant in the VR setting. The findings can provide useful guidance when processing 3D shapes for VR applications. In future work, we aim to construct a larger-scale database of perceptual quality under different combinations of distortions and build an objective quality assessment model to predict visual quality in VR.



(a) (b) (c)

Figure 5: An example of the distortion types. (a) Original Venus model and illustration of the different types of regions; (b) high-level noise applied on rough regions; (c) medium-level noise applied on smooth regions.

REFERENCES

[1] Pinar Akyazi and Touradj Ebrahimi. 2018. A new objective metric to predict image quality using deep neural networks. In *Applications of Digital Image Processing XLI*, Vol. 10752. SPIE, 554–567.

[2] Joao Ascenso, Pinar Akyazi, Fernando Pereira, and Touradj Ebrahimi. 2020. Learning-based image coding: early solutions reviewing and subjective quality evaluation. In *Optics, Photonics and Digital Technologies for Imaging Applications VI*, Vol. 11353. SPIE, 164–176.

[3] Abdullah Bulbul, Tolga Capin, Guillaume Lavoué, and Marius Preda. 2011. Assessing visual quality of 3-D polygonal models. *IEEE Signal Processing Magazine* 28, 6 (2011), 80–90.

[4] Meixu Chen, Yize Jin, Todd Goodall, Xiangxu Yu, and Alan Conrad Bovik. 2019. Study of 3D virtual reality picture quality. *IEEE Journal of Selected Topics in Signal Processing* 14, 1 (2019), 89–102.

[5] Zhengxue Cheng, Pinar Akyazi, Heming Sun, Jiro Katto, and Touradj Ebrahimi. 2019. Perceptual quality study on deep learning based image compression. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 719–723.

[6] Aladine Chetouani. 2017. A 3D mesh quality metric based on features fusion. *Electronic Imaging* 2017, 20 (2017), 4–8.

[7] Massimiliano Corsini, Elisa Drelie Gelasca, Touradj Ebrahimi, and Mauro Barni. 2007. Watermarked 3-D mesh quality assessment. *IEEE Transactions on Multimedia* 9, 2 (2007), 247–256.

[8] Massimiliano Corsini, Mohamed-Chaker Larabi, Guillaume Lavoué, Oldřich Petřík, Libor Váša, and Kai Wang. 2013. Perceptual metrics for static and dynamic triangle meshes. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 101–125.

[9] Jinjiang Guo, Vincent Vidal, Irene Cheng, Anup Basu, Atilla Baskurt, and Guillaume Lavoue. 2016. Subjective and objective visual quality assessment of textured 3D meshes. *ACM Transactions on Applied Perception (TAP)* 14, 2 (2016), 1–20.

[10] Jining Han, Qiyu Zheng, and Yunan Ding. 2021. Lost in Virtual Reality? Cognitive Load in High Immersive VR Environments. *Journal of Advances in Information Technology Vol* 12, 4 (2021).

[11] Joseph J LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin* 32, 1 (2000), 47–56.

[12] Guillaume Lavoué. 2009. A local roughness measure for 3D meshes and its application to visual masking. *ACM Transactions on Applied perception (TAP)* 5, 4 (2009), 1–23.

[13] Guillaume Lavoué, Irene Cheng, and Anup Basu. 2013. Perceptual quality metrics for 3D meshes: towards an optimal multi-attribute computational model. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 3271–3276.

[14] Guillaume Lavoué, Elisa Drelie Gelasca, Florent Dupont, Atilla Baskurt, and Touradj Ebrahimi. 2006. Perceptually driven 3D distance metrics with application to watermarking. In *Applications of Digital Image Processing XXIX*, Vol. 6312. International Society for Optics and Photonics, 63120L.

[15] Guillaume Lavoué and Rafał Mantiuk. 2015. Quality assessment in computer graphics. In *Visual signal quality assessment*. Springer, 243–286.

[16] Oliver A Meyer, Magnus K Omdahl, and Guido Makransky. 2019. Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education* 140 (2019), 103603.

[17] AA M Muzahid, Wanggen Wan, and Xiang Feng. 2018. Perceptual quality evaluation of 3d triangle mesh: a technical review. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. IEEE, 266–272.

[18] Yana Nehmé, Jean-Philippe Farrugia, Florent Dupont, Patrick Le Callet, and Guillaume Lavoué. 2020. Comparison of subjective methods for quality assessment of 3D graphics in virtual reality. *ACM Transactions on Applied Perception (TAP)* 18, 1 (2020), 1–23.

[19] BT Series. 2012. Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT* (2012), 500–13.

[20] Bernhard Spanlang, Jean-Marie Normand, David Borland, Konstantina Kilteni, Elias Giannopoulos, Ausiàs Pomés, Mar González-Franco, Daniel Perez-Marcos, Jorge Arroyo-Palacios, Xavi Navarro Muncunill, et al. 2014. How to build an embodiment lab: achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI* 1 (2014), 9.

[21] Gabriel Taubin. 1995. A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 351–358.

[22] Michela Testolina, Evgeniy Upenik, João Ascenso, Fernando Pereira, and Touradj Ebrahimi. 2021. Performance Evaluation of Objective Image Quality Metrics on Conventional and Learning-Based Compression Artifacts. In *13th International Conference on Quality of Multimedia Experience*.

[23] Fakhri Torkhani, Kai Wang, and Jean-Marc Chassery. 2014. A curvature-tensor-based perceptual quality metric for 3D triangular meshes. *Machine Graphics & Vision* 23, 1 (2014), 59–82.

[24] Libor Váša and Jan Rus. 2012. Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1715–1724.

[25] Kai Wang, Fakhri Torkhani, and Annick Montanvert. 2012. A fast roughness-based approach to the assessment of 3D mesh visual quality. *Computers & Graphics* 36, 7 (2012), 808–818.

[26] Zhou Wang and Alan C Bovik. 2011. Reduced-and no-reference image quality assessment. *IEEE Signal Processing Magazine* 28, 6 (2011), 29–40.

[27] Benjamin Watson, Alinda Friedman, and Aaron McGaffey. 2001. Measuring and predicting visual fidelity. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 213–220.

[28] Stefan Winkler. 2012. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing* 6, 6 (2012), 616–625.

[29] Rong Zhu and Yong Wang. 2012. Design and realization of virtual classroom. *Journal of Advances in Information Technology* 3, 1 (2012), 24–28.