

Fidelity Perception of 3D Models on the Web

Hussein Bakri^[0000-0001-9406-0116], Alan Miller^[0000-0003-1209-9063], and Iain Oliver

School of Computer Science, University of St Andrews, United Kingdom
{hb,alan.miller,iao}@st-andrews.ac.uk
<http://www.cs.st-andrews.ac.uk/>

Abstract Cultural heritage artefacts act as a gateway helping people learn about their social traditions and history. However, preserving these artefacts faces many difficulties, including potential destruction or damage from global warming, wars and conflicts, and degradation from day-to-day use. In addition, artefacts can only be present in one place at a time, and many of them can not be exhibited due to the limited physical space of museums. The digital domain offers opportunities to capture and represent the form and texture of these artefacts and to overcome the previously mentioned constraints by allowing people to access and interact with them on multiple platforms (mobile devices, tablets and personal computers) and network regimes. Through two experiments we study the subjective perception of the fidelity of 3D models in web browsers in order to discover perceptible resolution thresholds. This helps us create models of reasonable graphical complexity that could be fetched on the biggest range of end devices. It also enables us to design systems which efficiently optimise the user experience by adapting their behaviour based upon user perception, model characteristics and digital infrastructure.

Keywords: 3D Web · Heritage · Perception · Fidelity · Quality of Experience.

1 Introduction

In this work, we investigate the subjective perception of the fidelity (i.e resolution) of 3D Digital Heritage Artefacts on the web and how this affects the user experience. The models are hosted in the social repository Sketchfab (www.sketchfab.com).

The perception of fidelity is considered in the domain of Quality of Experience (QoE) which studies "the degree of delight or annoyance of the user of an application or service. It results from the fulfilment 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" [5].

By studying the perception of fidelity of digital heritage artefacts on the web, we aim to measure the differences in terms of fidelity across different categories of graphical complexity (i.e. resolutions). We therefore investigate if there are any noticeable differences detected by users between these categories and to what degree and at which resolution thresholds or ranges those differences become either *unnoticeable* or *intolerable*.

As anchor points for our study, we are interested in investigating the following:

1. The resolution limit or range that a device hardware & software can support.
2. The resolution limit or range below which, the fidelity of the model becomes unacceptable. We will call such limit the *lower resolution threshold*.
3. The resolution limit or range above which, users will notice little difference in fidelity. We will call such limit the *upper resolution threshold*.

This contribution of developing our understanding of the perception of fidelity is of great value in informing the design of systems that deliver optimised quality of experience within specific quality of service contexts.

We aim to discover if hard limits (i.e. thresholds) actually exist or if they fall into a range of values. Creating 3D models in the range of acceptable resolutions (between lower and upper thresholds) allows us to fetch to client devices lower levels of acceptable quality of 3D models thus improving the Quality of Service (QoS) mainly in terms of better download and processing time and better responsiveness especially on mobile devices. In addition, this helps us not to overcommit hardware resources that do not add much to the user experience hence achieving a trade-off between the fidelity of the 3D model and its responsiveness knowing that the higher the fidelity of the model, the lower the degree of its responsiveness and performance.

There is no need once the upper resolution threshold or range is discovered to fetch higher resolutions than that. In addition, we want to see if there is indeed a one size fits all solution, in other words, if there is a range of acceptable resolutions that we can fetch on the majority of end devices.

The results of these experiments have helped us design Hannibal, a QoS & QoE adaptive framework for virtual museums that aims to strike the best balance between QoS and QoE taking into consideration the capability of client devices and of network conditions.

The main contribution of this paper is the investigation of subjective perception of fidelity of 3D cultural heritage models on the Web with the aim of discovering resolution thresholds at which the fidelity of the models becomes either *unnoticeable* or *unacceptable*. This information may be of particular interest to curators of digital material in museums.

The remainder of this paper is organized as follows: Section 2 provides a background discussion on the importance of 3D digital artefacts in web-based virtual museums. Section 3 reviews related work in the domain of perception of fidelity of 3D models. Section 4 describes the design and experimental methodology of Experiment 1 (PC tethered to a big screen) and Experiment 2 (iPad Pro and iPhone 7 Plus) including a description of the apparatus used and the characteristics of Sketchfab 3D models. Section 5 presents the results from the two experiments and analyses them. Section 6 presents briefly the Hannibal adaptive Framework, which has been designed on the basis of the findings of this study, while Section 7 concludes.

2 Digital Heritage Artefacts and Virtual Museums

According to Shweibenz [18], a virtual museum is "a logically related collection of digital objects composed in a variety of media" which "lends itself to transcending traditional methods of communicating and interacting with the visitors being flexible toward their needs and interests" and "can be disseminated all over the world". In a sense, we could say a virtual museum is a "museum without walls" [3].

Traditionally, digital virtual museums presented cultural heritage artefacts on the web in multimedia forms such as audio commentaries, educational videos and images. In comparison to traditional media, 3D models provide a more interactive and richer user experience.

In the past, digitization required specialist equipments and expertise, but nowadays it is as easy as using commodity phones, cameras and freely available software. Digital heritage artefacts, sculptures and buildings are captured and digitally reconstructed through techniques like Photogrammetry [16], 3D scanners [4] often called laser scanners and topographical techniques [13].

It is also now possible to host these artefacts and disseminate them more easily on the web than it was possible in the past thanks to 3D on-line social repositories like

Sketchfab thus leading to their democratization for any web usage or audience and more specifically in the context of cultural heritage.

Sketchfab is a relatively new type of single media social archive that stores, embeds and renders 3D models on the web. Users can upload and comment on 3D models, like them, share them on their personal websites and social media sites and sell and buy them. Sketchfab is used extensively in cultural heritage applications like virtual museums on the web. It uses a 3D viewer based on the open source WebGL OSG.JS framework [1] which is an implementation of OpenSceneGraph concepts.

There has been a considerable delay in the emergence of 3D content on web platforms in terms of delivery and visualization compared to traditional digital media such images and videos and this is due to higher requirements of 3D graphics in terms of computation powers of client devices [8].

This began to change in recent years as we witnessed a steady and continuous proliferation of processing power, graphical capabilities of personal computers and mobile devices and a massive increase in available bandwidth. In addition, we witnessed the development of many 3D Web technologies like X3D, Oak3D and WebGL [2], taking advantage of the increased graphical capabilities of client devices.

The attraction of the web for delivering 3D objects and worlds is the nature of the web itself, which is very "accessible" to ordinary technology users. In addition, with the case of 3D material on the Web (3D Web), the same content might be displayed on a wide range of devices and environments. This of course means the display of 3D content faces the same issues of Quality of Service/Quality of Experience as other multimedia, mainly bandwidth, computational capabilities of end devices and content fidelity [22].

Most existing 3D Web systems suffer from latency (data download time) and lack of adaptation of level of details of 3D content to different transmission networks and hardware and software capabilities of end devices [10]. Virtual museums are no exception hence in this work, we are interested in investigating subjective perception of fidelity (i.e level of detail) thresholds with the aim of achieving a trade-off between fidelity vs performance of heritage artefact models, and thus enhancing the overall performance and user experience of virtual museums. The following section presents examples of previous research that assessed fidelity perception in 3D graphics and multimedia.

3 Assessing perception of fidelity of 3D Models

Much research has been conducted on assessing the subjective perception of 3D static and dynamic models, but the majority of these research studies did not address 3D web context nor the study of fidelity thresholds. Although some of them have used 3D Web tools in their subjective tests, the usage of such tools was not an aim in itself. The majority of cases involved measuring the effect of a certain graphical algorithm on distorted 3D meshes or texture maps shown usually as videos or animations.

The subjective perceptual quality of textured 3D Mapped models was evaluated by Guo et al. [9] using a paired comparison protocol (2 stimuli - side by side) on videos of 136 texture- and geometry-distorted models of 5 reference models on a total of 101 subjects. Participants were asked to choose the model decimation that is similar to the reference model. They also proposed two new objective metrics for visual quality assessment of textured meshes. Although their interface was developed as a web platform using JavaScript, the platform was used only to show models as videos.

The effect of lighting environment on the perception of fidelity of 3D models was measured by Vanhoey et al. [20], who conducted a perceptual study on 36 participants. The models were presented to the participants as video sets. They used 2 stimuli side-by-side paired comparison protocol. The problem of assessing distortions produced

by watermarking 3D meshes was studied by Corsini et al. [7]. Pan et al. [12] studied the importance of the relation of wireframe to texture resolution in the perception of quality of 3D Objects. Animations were used as a user interaction method in their study. Watson et al. [21] and Rogwitz et al. [17] focused on studying simplification algorithms using double stimulus rating scenarios where the subjects rate the fidelity of simplified models in comparison with original ones.

In the field of multimedia, much research has been conducted. For example, Talens-Noguera et al. [19] conducted an experiment on users to measure the perception of quality variation over time (QVT) over a set of videos. They used a single stimulus ITU-R Absolute Category Rating (ACR) 5-point scale, a system similar to the one we used in our experiments.

This cursory review of related works suggests that the question of investigating the upper and lower resolution thresholds for the 3D web models in a digital heritage context is an under-researched area worthy of study.

4 Experimental Methodology

In order to obtain representative ratings, we followed recommendations from the multimedia discipline in choosing the number of participants in quality perception experiments. According to the recommendation "P910 - Subjective video quality assessment methods for multimedia applications" [15] which is authored by the Video Quality Experts Group (VQEG), the number should be above 4 for statistical reasons and below 40. There is no added value of having more than 40 participants according to the recommendation.

In Experiment 1, 22 participants were recruited. This comprised 7 females, 14 males and 1 classified as other; between the age of 22 and 50, all with normal or corrected to normal vision. The mean age of participants was $\mu = 30$ and the standard deviation was $\sigma = 5.9$. In Experiment 2 (on a tablet and a mobile phone), 10 participants were recruited. This comprised of 3 females and 7 males between the age of 21 and 40, all with normal or corrected to normal vision. The mean age of participants was $\mu = 27.4$ and the standard deviation was $\sigma = 5.04$. Very few of the participants in both experiments were really involved in dealing with 3D models in particular and computer graphics in general.

Some of the participants in Experiment 2 were the same participants in Experiment 1. Learning effect is not present due to the fact that all models were randomized in both the grading and ranking phases of both experiments. In addition, experiment 1 and 2 were done separately from each others (approximately more than 2 weeks apart). The participants were presented with a questionnaire concerning basic demographic information (age and gender) in addition to information about eye sight. This questionnaire included also information about regular viewing preferences of participants for 3D content and 3D modelling experience. Participants were sat at a distance of 2 meters from the big screen (i.e a distance of around $2.5H$, where H is the height of the display $H = 80\text{cm}$).

4.1 Reference Models

The reference models were reconstructed from real world digital heritage artefacts. The Achavanich beaker was digitized through Photogrammetry. A statuette of the Roman god, Mercury, was digitally scanned using a 3D laser scanner.

Roman Mercury Model: depicting a bronze statuette of the god. The model is shown in figure 1. The model native resolution is 1.5M faces, and 733.5K Vertices. Its size on disk is 38MB. This model was chosen because of its simple and loosely defined



Figure 1. Mercury Model



Figure 2. Achavanich Beaker Model

topological features and its darker textures. The choice would help us in understanding the effect of those features.

Achavanich Beaker Model: depicting the Achavanich Beaker which was found buried with the remains of a 4,000 year old Bronze Age woman, Ava, in the Northern Highlands of Scotland. The model is shown in Figure 2. The model initial resolution is 6.5M faces, and 3M Vertices. Its size on disk is 116MB. Achavanich Beaker is a model with brighter textures and more complex topological features than that of Mercury model.

Decimation is the process of reducing the number of faces of a 3D model’s mesh also known as *mesh simplification*. The models were decimated into lower resolutions using the Quadric Edge Collapse Decimation algorithm preserving UV parametrizations [14]. A filter Meshlab (meshlab.sourceforge.net) application which implements the algorithm was used. Textures were unaffected.

Seven decimation levels were used for the **Roman Mercury** statue in Experiment 1: [50K, 100K, 183K, 300K, 366K, 733K, Original (1.5M)] and ten for the **Achavanich Beaker**: [100K, 150K, 189K, 300K, 375K, 750K, 1M, 1.5M, 3M, Original resolution (6.1M)].

In Experiment 2, higher fidelity models were excluded as they would not run on the hardware. The following decimations were used: **Roman Mercury**: 7 decimated [50K, 100K, 183K, 300K, 366K, 733K, Original resolution (1.5M)]. These are the same resolutions used in Experiment 1 **Achavanich Beaker**: 7 decimated models were used [100K, 150K, 189K, 300K, 375K, 750K, 1M]. The highest resolution that can be rendered without crashes on both mobile devices used is the 1M resolution.

The choice of resolutions was made in such a way as to make comparisons across models and devices but this was sometimes not possible since it was more dictated by how the decimation algorithm works when minimizing the number of faces while still preserving the integrity of the 3D meshes.

4.2 Client Devices Specifications

In **Experiment 1** The system that rendered the models was an Intel Core i5-440-3.10 GHz with 16GB 1067 MHz DDR3 RAM. The graphics card of the machine was NVIDIA GeForce GTX 970 with 4GB of Video RAM. The system has an installation of Windows 7 Enterprise 64 Bit with a minimal set of background processes running to avoid interference. The system had the latest drivers installed and the web browser used was Google Chrome version 57.0.2987.98 (64-bit). The screen used to display the models was a Sharp Screen 60-inch Full HD with 1920x1080 native resolution. The system was connected to the screen with an HDMI 2.0 connection. The use of relatively powerful consumer hardware is necessary to render the complex 3D models with millions of faces and vertices. The network connection of the system is to a 100 Mbps Ethernet Local

Table 1. Five level quality of presentation scale

5-level Score Subjective Assessment	
5	Excellent Resolution
4	Good Resolution
3	Fair Resolution
2	Poor Resolution
1	Bad Resolution

Area Network which connected by four 1Gpbs local hops to the UK Joint Academic Network (JANET) backbone.

In **Experiment 2**, Two devices were chosen that pertain to two categories (smart phones and tablets) in this experiment:

iPad Pro (9.7 inch) WiFi 32GB MLMN2B/A model: has as a processor the Apple A9X (64 bit architecture). The GPU inside the A9X chip is the PowerVR Series 7XT GT7600 (six-core). The Operating System (OS) on the iPad Pro was iOS 10.3(14E277). The mobile web browser used was Opera Mini version 14.0.0.104835 (64-bit). We found that Opera Mini was the best WebGL browser on the iPad compared to other 4 common mobile browsers benchmarked for the purpose of the experiment (Safari, Firefox, Google Chrome and Dolphin). The experiments were conducted on this device, with a "portrait" screen orientation.

iPhone 7 Plus 256 GB: has a 5.5-inch LED-backlit wide screen with Retina HD Display technology. It has as a processor, the A10 Fusion chip 64-bit architecture (Embedded M10 motion coprocessor). The GPU inside this chip is the PowerVR Series 7XT GT7600 Plus (hexa-core). The OS version on the iPhone 7 Plus was iOS 10.3-14E277. The experiments were conducted on this device with a "portrait" screen orientation. The mobile web browser used on this device was Safari (64-bit). We found that Safari was the best WebGL browser on the iPhone 7 Plus compared to other 4 common mobile browsers benchmarked for the sake of the experiment (Opera Mini, Firefox, Google Chrome and Dolphin). The iPhone 7 Plus was considerably better in terms of graphical processing power than the aforementioned iPad Pro. This was shown by the fact that it was able to render a resolution even higher than the 1M faces, which is the highest that iPad Pro can render reliably.

Both mobiles devices were connected to a stable WiFi network connected to Access Points that support data rates of around 600Mbps.

4.3 Experimental procedure

The software used to capture participants' responses was a PHP web application connected to a MySQL database which rendered the models in different views and resolutions, took the questionnaire answers from participants and stored them in several tables in the database for later statistical analysis. The same web application was used with relatively minor changes for the experiment conducted on the iPad Pro and the iPhone 7 Plus.

The participants were made aware that they were grading and ranking only the quality of presentation or fidelity of the 3D models and not responsiveness of the 3D model or qualities of the original artefact. The "reference models" and all their "decimations" were rendered using the "**default rendering**" settings in Sketchfab on a black background. Each experiment was conducted into grading and ranking phases. For grading we used a single stimulus ACR-HR (Absolute Category Rating with Hidden



Figure 3. 6.1M Resolution **Figure 4.** 300K Resolution **Figure 5.** 100K Resolution

Reference) protocol similar to what is recommended in the P910 [15]. Participants were asked to grade the models, one at a time, based on 5-level scale presented in table 1. Model variations were shuffled randomly on different pages and graded on a scale of 1 to 5 for user perceived fidelity. Subjects graded each model twice first without interaction and then with interaction.

Participants then ranked models from worst to best perceived fidelity. No clues were provided concerning the model resolution. Five 3D model decimations labelled A to E were present on the same web page where participants could interact with any model resolution they wished to. The aim was to identify resolutions above which participants began to confuse the correct ranks.

We asked the participants to grade the models based on a 5-level quality of presentation scale shown in Table 1. The rationale behind that is that it is a typical scale widely used in the perception of fidelity of multimedia [19] and 3D models [12] and recommended by P910 [15]. We considered using a *9-level quality scale* but decided that more grades to describe same 5 categories were unnecessary.

In the grading phase, we calculated the Mean Opinion Score (MOS) for each resolution, which is defined as follows:

$$MOS = \frac{\sum_{n=0}^N R_n}{N} \quad (1)$$

Where N is the number of participants, R_n is all the ratings given by all participants of a certain stimulus (as an example: the 100000 faces resolution of Achavanich beaker model in no interaction mode on a 60 inches HD screen).

In the ranking phase, we calculated the Absolute Error which is defined as follows:

$$\Delta x = x_0 - x \quad (2)$$

Where Δx is the absolute error of a particular ranking response, x_0 is the response of the participant (i.e. the rank given by the participant) and x is the actual theoretical rank of a certain resolution. We also calculated the Mean Absolute Error (MAE) of all resolutions. The following section describes the results of the QoE experiments.

5 Grading and Ranking Results and Analysis

This section presents the results obtained from users grading and ranking the decimated models, on a computer, iPad Pro and iPhone 7 Plus. We discuss the discovery of thresholds and compare results from passive observation and interaction.

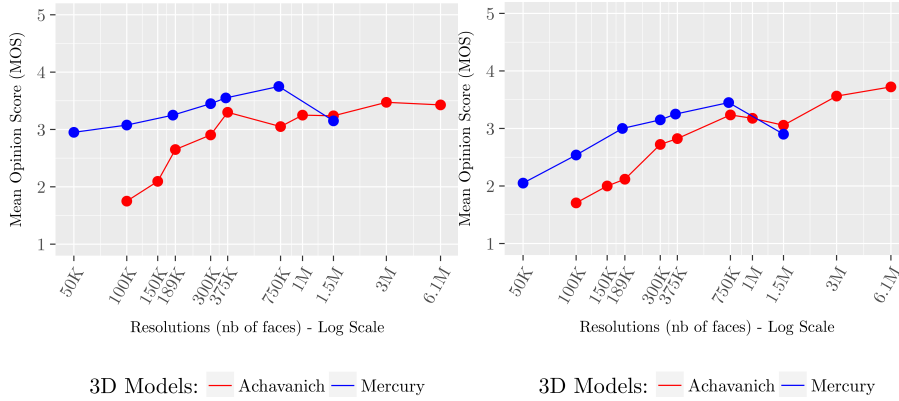


Figure 6. Big Screen - No Interaction **Figure 7.** Big Screen - With Interaction

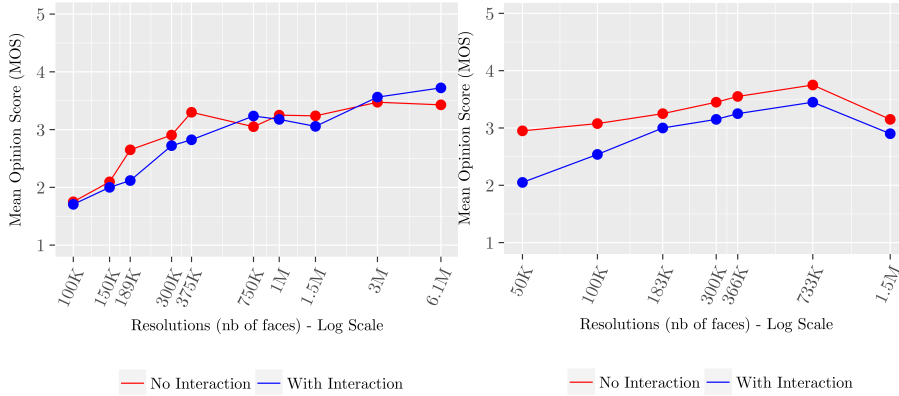


Figure 8. Interaction - Achavanich

Figure 9. Interaction - Mercury

5.1 Hardware & Software resolution Limits

Hardware limits: The iPhone 7 Plus and the iPad Pro were able to fetch successfully the 1.5M resolution (i.e. original resolution) of the Mercury Model, which is a simple model with loosely defined topological features and darker textures. The devices could not fetch successfully more than the 1M faces as a resolution in Achavanich Beaker, a model with brighter textures and more complex topological features than that of Mercury model. The PC with high-end graphics card was able to fetch all the resolution decimations of the 2 models.

Software limits: We benchmarked 5 web browsers on the devices used (PC, iPad Pro and iPhone 7 Plus). The web browsers were Google Chrome, Opera (for PC), Opera mini (for mobile devices), Mozilla Firefox, Apple Safari, the Dolphin Browser (for mobile devices).

We discovered that web browsers play an important role in terms of software resolution limits on each device as some browsers were able to fetch higher resolutions on mobile devices than other browsers (this is probably due to different implementations and configurations of WebGL). For example, on the iPad Pro, the Opera Mini browser was able to render reliably the 1M resolution of the Achavanich Beaker, whereas for the same resolution, on the same device, Safari crashed many times. The best WebGL browser for the PC running Windows was Google Chrome, for the iPad Pro was Opera mini and for the iPhone 7 Plus was Apple Safari.

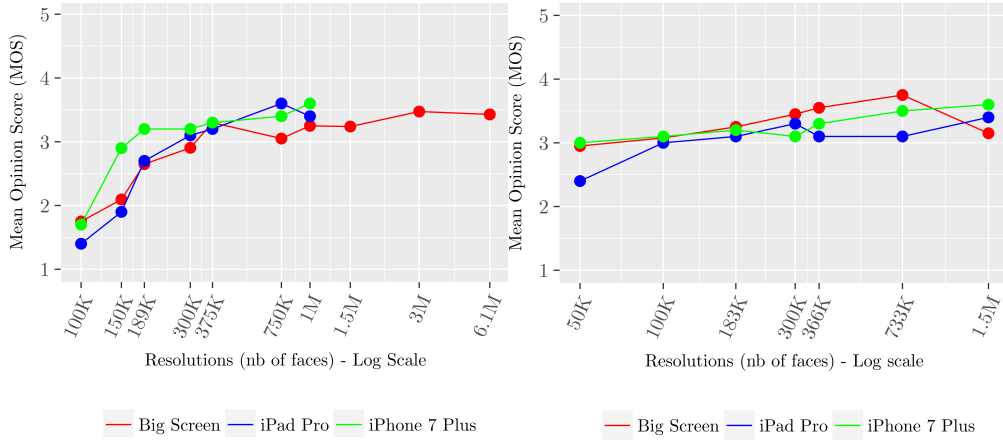


Figure 10. Achavanich - No Interaction

Figure 11. Mercury - No Interaction

5.2 Upper and Lower Resolution Thresholds

The *Upper Resolution Threshold range or limit* of a 3D model is the resolution range or limit above which any improvement in fidelity is not noticeable or "barely" noticeable by users. In some cases, any improvement of fidelity above the upper threshold range could even be detrimental to users' overall perception of the quality of presentation for several reasons. Mainly the 3D model becomes "laggy" but more importantly it will have a new redistribution of the normals and vertices clouds which could have a reverse effect on the overall perceived fidelity. After such a threshold range, users might judge a lower resolution as better than a higher resolution.

The Lower Resolution Threshold or range of a 3D model is the resolution limit or range below which fidelity is unacceptable to users. This was difficult to discern from the results of our experiments as this limit, if it exists, is more a judgement call of participants on what is considered as an unacceptable or intolerable quality of presentation.

Results in Figure 6 and Figure 7 present a comparison between the grades of the 2 models, the Achavanich Beaker and Mercury, on a 60 inch HD screen with and without interaction. Both show that the Mercury model is relatively a better model, scoring higher MOS values than the Achavanich Beaker. For instance, the 50K resolution in the Mercury model has an MOS of around 3 (fair resolution) whereas the 100K resolution in the Achavanich Beaker has a MOS of around 1.75.

We can observe from Figure 6, that *around the range of 300K to 375K, we cease to get any significant benefit in perceived fidelity for Achavanich Beaker model*. This is due to the fact that above this range, meaning between 375K and 6.1M (a big range of resolutions on a logarithmic scale), there is hardly any significant increase in MOS values (around only 0.25).

In a similar vein, for the Mercury model in no interaction mode, *the range between 300K and 375K* seems also to be the range of resolutions above which there is no significant or apparent benefit. This is due to the fact that there is *barely* any increase between 375K and 750K and after the 750K resolution, the MOS value decreased to around 3.

In *interaction mode, the range between 375K and 750K resolutions is the range above which we cease to get any significant benefit in the quality of presentation*. This is shown in Figure 7, the increase in MOS is only around 0.4 between 750K and 6.1M in Achavanich beaker model. Figure 8 shows the effect of interaction on Achavanich beaker.

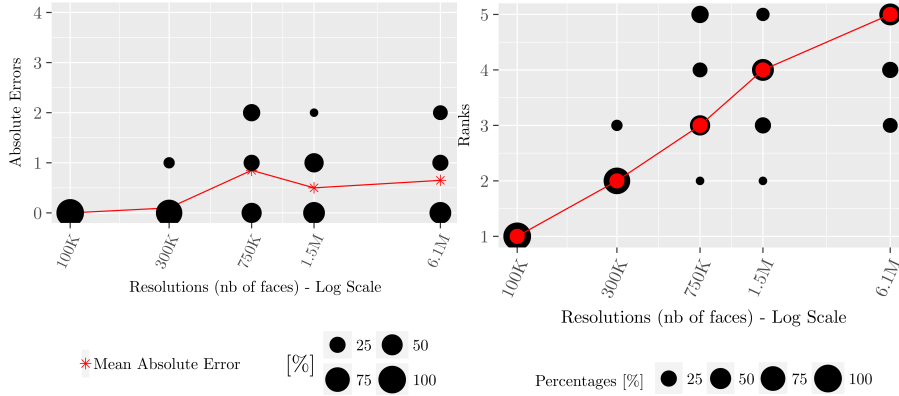


Figure 12. Absolute Errors (left) & Ranks Distributions (right) for Achavanich on Big Screen

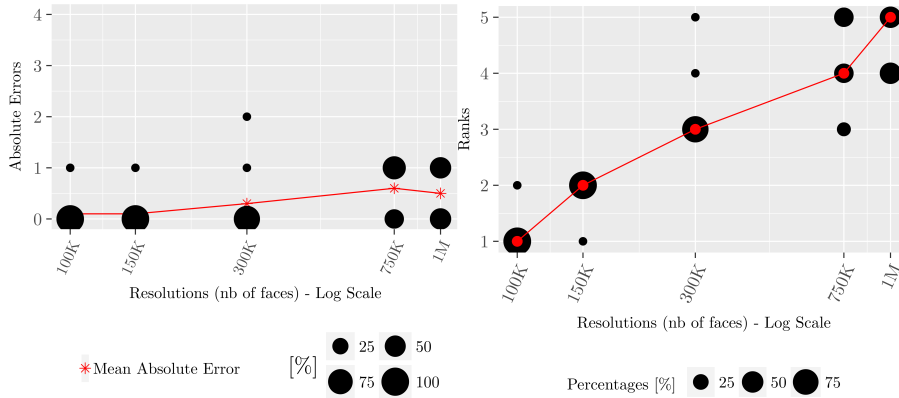


Figure 13. Absolute Errors (left) & Ranks Distributions (right) for Achavanich on iPad Pro

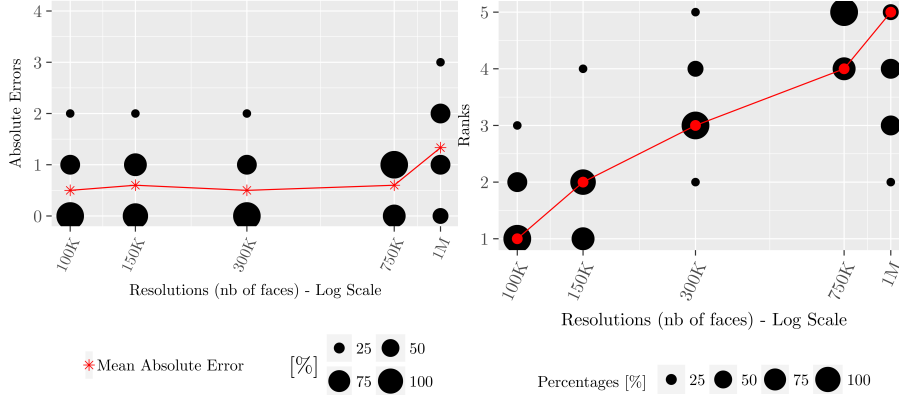


Figure 14. Absolute Errors (left) & Ranks Distributions (right) for Achavanich on iPhone

This was not apparent in Mercury model as the *upper resolution range remained with interaction between 300K and 375K*. This can be seen both in Figure 7 and Figure 9.

When resolutions are ranked, Figure 12 shows that for the Achavanich Beaker on a big screen, *participants begin to confuse ranking after the resolution of 300K* and that resolution 750K has a Mean Absolute Error (MAE) of around 1 with absolute errors distributed between 0, 1 and 2, meaning some participants ranked this model's resolution 2 positions off-setted from its correct position.

Results for **mobile devices** are shown in Figure 10 and Figure 11 show the variation of MOS values for the iPhone 7 Plus and iPad Pro in no interaction mode for both models. We can observe for Achavanich beaker model on iPhone 7 Plus (5.5 inches),

that after *the range of 150K to 189K*, we cease to get any significant benefit in the quality of presentation. This is due to the fact that higher than this range meaning between 189K and 1M (a big range of resolutions on a logarithmic scale), there is hardly any significant increase in MOS values. *For Achavanich Model on the iPad Pro (9.7 inches), the upper range is between 189K and 300K.* On the other hand, for *Mercury model almost same ranges for mobile devices (183K to 300K).*

When resolutions are ranked, Figure 13 shows for Achavanich Beaker on iPad Pro, *participants begin to confuse ranking after the resolution of 150K.* In addition, it shows that resolution 300K has absolute errors which are distributed between 0, 1 and 2, meaning some participants ranked this model 2 positions off-setted, whether for 150K, absolute errors were divided between 0 (no error) and 1. Concerning the ranking on the iPhone 7 Plus, Figure 14 shows that *for the 100K resolution*, we see absolute errors taking values of 0 (no error), 1 (rank offset by 1) and 2 (rank offset by 2). *This shows that it is more difficult to detect difference of fidelity for screens of 5.5 inches.*

5.3 Effect of Interaction

In this section, we study the effect of interaction on both models. Figure 8 shows the effect of interaction on Achavanich Model fetched on a 60" HD screen. With interaction for Achavanich Beaker model, lower resolutions have higher MOS values compared to when interaction is not allowed. In other words, interaction lowers the MOS value of a resolution. On the other hand, interaction seem not to have any apparent effect for Achavanich Beaker model for higher resolutions (above the 375K resolution). This is due to the fact that we see the 2 lines representing interaction modes fluctuates after the 375K while increasing.

Figure 9 shows the effect of interaction on Mercury Model fetched on a 60" HD screen. It seems clear that with interaction, all resolutions (lower and higher) have lower MOS values than their counterparts in no interaction mode. At higher resolutions in a brighter textured model, interaction (zooming in, zooming out, panning) seems to have *little effect on perception of fidelity.* This is not the case for a darker textured model like the Mercury model.

Interaction plays a big role in the perception of fidelity for lower resolutions where we can see in both Figure 8 and Figure 9 a big gap of MOS values in lower resolutions of each model. After the upper threshold range, interaction has little effect on perception of difference of fidelity. This is shown by the smaller gap of MOS between the 2 lines representing the interaction and non-interaction modes.

5.4 Effect of Screen Sizes

In this section, we investigate if screen size have any effect on the perception of fidelity. Figure 10 shows variation of MOS values of Achavanich Beaker resolutions in no interaction mode, across devices (PC tethered to big screen, iPad Pro and iPhone 7 Plus). Figure 11 shows variation of MOS values of Mercury resolutions in no interaction mode across all devices (big screen, iPad Pro and iPhone 7 Plus).

It is not conclusive that screen size in our experimental setting has any discernible effect. As for the Achavanich Beaker model, the big screen seems to have lower MOS values than mobile devices, while for the Mercury model the opposite is apparent for a large number of resolutions. This may be due to the fact that having participants sit 2 meters away from a very large HD screen (60 inches) seems to have the same effect as letting people interact with models on screens of 9.7 inch for iPad Pro and 5.5 inch for iPhone 7 Plus. In addition, two parameters may be involved here: the resolution of the screen in question and its size. The iPhone 7 Plus has a 4K fidelity but a small screen size.

5.5 Further measurement work

There are several areas where this measurement work can be further developed. For example, additional 3D models with different categorized topologies would give a more comprehensive view. The two models chosen were obtained from digitizing actual historical heritage artefacts through Photogrammetry and 3D Scanning. They might be different from models constructed in 3D authoring applications like Blender or Maya as "photogrammetised" or scanned models tend to have more complex texture seams. There is also the need to address the effect of texture resolutions, types and compression on the subjective perception of fidelity of 3D models on the web and what the thresholds are then in this regard. This is the subject of an empirical investigation we are currently conducting. It is pertinent to confirm and buttress results with a bigger population especially in Experiment 2 (on iPad Pro and iPhone 7 Plus).

The experiments gave us important insights into how a dark textured and loosely defined model in terms of surface topology and features like the Mercury model would be graded and ranked. We also, investigated the same for a brightly textured model with a more complex topology (Achavanich Beaker). Latency affects the QoS and thus the overall QoE, but here we focus on the client performance, consequently models were pre downloaded and rendered in the web browsers.

6 Hannibal - An Adaptive Framework for Virtual Museums

The most important takeaway benefit from the results of the experiments presented in this paper in addition to other results from experiments tackling Quality of Service (QoS) metrics that we also captured in another as yet unpublished work, is to help us in the design and implementation of Hannibal, a QoS & QoE aware adaptive framework for virtual museums.

In a virtual museum, choreographing and presenting 3D Web components, mainly 3D models, across multiple platforms (mobile devices, tablets and personal computers) and network regimes (5G, 4G, 3G or WiFi and Broadband Ethernet) with varying bandwidth and latency present a significant challenge yet to be overcome.

Therefore, this makes it difficult to achieve a good user QoE across all these platforms, which have different characteristics and capabilities. This means that different levels of fidelity and complexity of media may be appropriate and should be available. Thus servers hosting those media types need to adapt to the capabilities of a wide range of networks and client devices.

Hannibal is named after "Hannibal Barca", the hero of the Phoenician Carthaginians who was himself adaptive and cunning in his war tactics against the Romans during the Punic wars. Hannibal aims to allow virtual museum applications to deliver the best possible user experience across platforms. The engine will be integrated in the Virtual Museum Infrastructure (VMI) [6], a complete web-based virtual museum infrastructure.

3D models are uploaded through a management interface to Omeka (www.omeka.org), an asset management system. These models are decimated to lower specific fixed fidelities. The idea is inspired from YouTube Adaptive System [11], which decimates an uploaded video into segments of specific lower bit rates or resolutions.

QoS-related metadata like numbers of faces and number of vertices among others are also stored with the 3D models. This helps Hannibal via a decision expert system to first detect the client capability and network conditions and then to fetch the right resolution decimation that is convenient to the *situation of the client device* (hardware and software resolution thresholds and network conditions). Hannibal also takes into consideration the user preference from a user interface in terms of trade-off between Fidelity of models Vs Performance.

Table 2. Hardware Resolution Limits of both models

3D Models	iphone 7 Plus	iPad Pro	PC
Achavanich Beaker Model	1M	1M	6.1M
Mercury god Model	1.5M	1.5M	1.5M

7 Conclusions

In this work, we investigated the subjective perception of the fidelity (i.e resolution) of 3D digital heritage artefacts in web browsers and how this affects the user experience. The models were digitized from actual historical heritage artefacts via Photogrammetry and 3D scanning techniques. They are hosted on the social repository Sketchfab. We divided this investigation into two user study experiments. The first experiment were conducted on a PC with a powerful graphics card tethered to a 60" screen. The second experiment was conducted on an iPad Pro tablet and on an iPhone 7 Plus.

We found that as resolution increases above a threshold it becomes harder for users to perceive a benefit. This upper resolution range is discernible. For a big screen, it is between 300K and 375K. For an iPad Pro (9.7 inches screen), this range becomes roughly between 189k and 300K. For an iPhone 7 Plus, this range is between 150K to 189K. These findings are supported by the results that we obtained from the ranking phase. This suggests that there is a sweet spot where a wide range of devices can be supported by using models of around 300K faces.

We discovered, from our experiments, that each device has a hardware and software limit of fetching certain resolutions. This is important when it comes to mobile devices with limited graphical capabilities. Table 2 presents the hardware resolution limits of the 3 devices used in the experiment (PC with Big Screen, an iPad Pro and iPhone 7 Plus). It shows clearly that mobile devices (tablet and phone) were not capable of fetching all the range of resolutions that a PC with a high-end graphics card can render.

We discovered that screen size has marginal effect but that with small screens like those of mobile and tablet devices, users tolerate lower resolutions because they could not tell the difference between many of them. There is not a hard threshold on the lowest resolution of value rather perception continues to degrade as resolution is reduced.

Interaction lowers the perception of fidelity of the models as people zoom in and out and thus discover imperfections in the models. We discovered that this effect (Interaction effect) is most apparent at lower resolutions below the upper range threshold than at higher resolutions. At high resolutions, interaction has less effect on the perception of the overall quality of presentation of a model. However, where close inspection of model detail is important to the user activity there is motivation for delivering a high resolution model provided the client has the appropriate capabilities

We have learnt that for 3D models some devices do not offer effective support for 3D, for those that do most devices and activities can be supported ensuring models within the sweet spot of around 300K, that the sweet spot may be lower depending on device and model and that for certain user activities higher resolution models will be of benefit. This suggests that the design of a framework for optimising the user experience base upon client characteristics and user experience is both a practical proposition and will yield substantial benefits to its users.

Acknowledgements

We express our gratitude to Jane Brooks and Mary Carr from ELT, University of St Andrews, for helping out in the proof reading and language of this paper.

References

1. OSG.JS WebGL Framework. <http://osgjs.org/>, accessed: 2018-03-01
2. Bakri, H., Allison, C., Miller, A., Oliver, I.: Virtual Worlds and the 3D Web—time for convergence? In: International Conference on Immersive Learning. pp. 29–42. Springer (2016)
3. Bearman, D.: Interactive multimedia in museums. *Studies in Multimedia: State-of-the-Art Solutions in Multimedia and Hypermedia* pp. 121–137 (1991)
4. Böhler, W., Marbs, A.: 3D scanning and photogrammetry for heritage recording: a comparison. In: Proceedings of the 12th International Conference on Geoinformatics. pp. 291–298. Gavle University Press, Sweden (2004)
5. Brunnström, K., Beker, S.A., De Moor, K., Dooms, A., Egger, S., Garcia, M.N., Hossfeld, T., Jumisko-Pyykkö, S., Keimel, C., Larabi, M.C., et al.: Qualinet white paper on definitions of quality of experience (2013)
6. Cassidy, C.A., Fabola, A.E., Miller, A.H.D., Oliver, I.A.: A digital museum infrastructure for preserving community collections from climate change. *iLRN 2017 Coimbra* (2017)
7. Corsini, M., Gelasca, E.D., Ebrahimi, T., Barni, M.: Watermarked 3-D mesh quality assessment. *IEEE Transactions on Multimedia* **9**(2), 247–256 (2007)
8. Di Benedetto, M., Ponchio, F., Ganovelli, F., Scopigno, R.: SpiderGL: a JavaScript 3D graphics library for next-generation WWW. In: Proceedings of the 15th International Conference on Web 3D Technology. pp. 165–174. ACM (2010)
9. Guo, J., Vidal, V., Cheng, I., Basu, A., Baskurt, A., Lavoue, G.: Subjective and Objective Visual Quality Assessment of Textured 3D Meshes. *ACM Transactions on Applied Perception (TAP)* **14**(2), 11 (2016)
10. Lavoué, G., Chevalier, L., Dupont, F.: Streaming compressed 3D data on the web using JavaScript and WebGL. In: Proceedings of the 18th international conference on 3D web technology. pp. 19–27. ACM (2013)
11. Lederer, S., Müller, C., Timmerer, C.: Dynamic adaptive streaming over HTTP dataset. In: Proceedings of the 3rd Multimedia Systems Conference. pp. 89–94. ACM (2012)
12. Pan, Y., Cheng, I., Basu, A.: Quality metric for approximating subjective evaluation of 3-D objects. *IEEE Transactions on Multimedia* **7**(2), 269–279 (2005)
13. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C.: Methods for 3D digitization of cultural heritage. *Journal of cultural heritage* **8**(1), 93–98 (2007)
14. Pietroni, N., Tarini, M., Cignoni, P.: Almost isometric mesh parameterization through abstract domains. *IEEE Transactions on Visualization and Computer Graphics* **16**(4), 621–635 (2010)
15. Rec, I.: P. 910: Subjective video quality assessment methods for multimedia applications. *Int. Telecomm. Union, Geneva* (2008)
16. Remondino, F.: Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sensing* **3**(6), 1104–1138 (2011)
17. Rogowitz, B.E., Rushmeier, H.E.: Are image quality metrics adequate to evaluate the quality of geometric objects? In: *Photonics West 2001-Electronic Imaging*. pp. 340–348. International Society for Optics and Photonics (2001)
18. Schweibenz, W.: The Virtual Museum: New Perspectives For Museums to Present Objects and Information Using the Internet as a Knowledge Base and Communication System. *ISI* **34**, 185–200 (1998)
19. Talens-Noguera, J.V., Liu, H.: Studying the Perceived Quality Variation over Time for Video Quality Assessment. In: Proceedings of the 1st International Workshop on Perception Inspired Video Processing. pp. 35–36. ACM (2014)
20. Vanhoey, K., Sauvage, B., Kraemer, P., Lavoué, G.: Visual Quality Assessment of 3D Models: On the Influence of Light-Material Interaction. *ACM Transactions on Applied Perception (TAP)* **15**(1), 5 (2017)
21. Watson, B., Friedman, A., McGaffey, A.: Measuring and predicting visual fidelity. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 213–220. ACM (2001)
22. Zampoglou, M., Kapetanakis, K., Stamoulias, A., Malamos, A.G., Panagiotakis, S.: Adaptive streaming of complex Web 3D scenes based on the MPEG-DASH standard. *Multimedia Tools and Applications* pp. 1–24