Subjective Evaluation of High-Fidelity Virtual Environments for Driving Simulations

Kurt Debattista*, Thomas Bashford-Rogers†*, Carlo Harvey‡*, Brian Waterfield§, Alan Chalmers*

*WMG, University of Warwick
†University of the West of England
‡Birmingham City University
§Jaguar Land Rover

Abstract—**Virtual environments grant the ability to experience real-world scenarios, such as driving, in a virtual, safe and reproducible context. However, in order to achieve their full potential the fidelity of the virtual environment must provide confidence that it replicates the perception of the real-world experience. The computational cost of simulating real world visuals accurately means that compromises to the fidelity of the visuals must be made. In this work a subjective evaluation of driving in a virtual environment at different quality settings is presented. Participants (n = 44) were driven around in the real world and in a purposely built representative virtual environment and the fidelity of the graphics and overall experience at low, medium and high visual settings were analysed. Low quality corresponds to the illumination in many current traditional simulators, medium to a higher quality using accurate shadows and reflections and high to the quality experienced in modern movies and simulations that require hours of computation. Results demonstrate that graphics quality affects the perceived fidelity of the visuals and the overall experience. When judging the overall experience, participants could tell the difference between the lower quality graphics and the rest but did not significantly discriminate between the medium and higher graphical settings. This indicates that future driving simulators should improve the quality, but once the equivalent of the presented medium quality is reached, they may not need to do so significantly.**

Index Terms—**virtual environments, virtual reality, driving simulators, subjective evaluation, physically-based rendering**

I. INTRODUCTION

Virtual environments (VEs) offer the possibility of investigating a real world scenario in a safe, controlled and repeatable manner [1]. While many applications do not need physically-based realism in order for users to accomplish a task in the VE, for others such accuracy is essential if the user is to have confidence that what is experienced in the virtual world is equivalent to what may be experienced in the real world. For example, inaccurate representation of glare on a screen may make the screen contents visible in the VE, when in fact, due to bright sunshine coming through the window, they are not visible in the real-life environment.

VEs are increasingly being used in the automotive industry to speed up development and save costs in vehicle design and manufacture by minimising the need to build physical prototypes and reducing time to market. In particular, driving simulators are frequently used to study the behaviour of drivers under distinct scenarios which would be difficult to replicate in the real world [2]–[4]. In order to fully exploit VEs, a high

level of confidence is needed to ensure that decisions taken in the virtual world are equivalent to those that would be taken in the real world scenario being considered. To authentically simulate reality, VEs need to be based on accurate physical simulations.

For lighting, such physical simulations require the precise modelling of objects, the reflectance models of light-to-surface interaction, and the high-fidelity simulation of the transport of light in an environment. Despite significant improvements in the performance of computer hardware it is still not possible to compute the physical simulation of light transport within a complex environment in real time, or even reasonable time. This limitation poses questions on whether complex physically-based simulation of lighting is in fact necessary for virtual driving scenarios or whether a lower quality, and thus computationally less demanding, simulation could still provide the necessary confidence.

This work investigates the perceived quality of the visuals in a driving scenario and how, in turn, this may affect the overall perception of the fidelity of a driving simulator. An experiment was undertaken in which participants experienced a real world driving scenario and were asked to judge the virtual equivalent of the real-world experience, in a practical driving simulator, see Figure 1. Three different levels of simulation accuracy for the visuals were used: a high-fidelity physically based simulation, a medium quality version, and a low quality simulation. The low quality simulation is comparable to that used to simulate light transport in interactive methods. The medium quality adds accurate visibility and introduces reflections. The high quality setting corresponds to high-fidelity simulations and adopts the same methods used in highly realistic simulations and for special effects and animated movies that require hours of rendering time to compute an image [5]. A virtual reconstruction of the real track was created by means of photogrammetry, see Figure 2, and the participants experienced the same scene under the different quality settings as passengers. Furthermore, authentic acoustic effects and motion were recorded and replayed in the practical driving simulator used in the experiment.

This is the first study of VEs for evaluating the fidelity of visuals in driving simulators. Furthermore, it allowed participants to experience the same environment in a real world track and virtual world track. The main contribution of this work is the finding that participants have a preference for

higher quality visuals, higher than what is typically used in modern driving simulators, although they do not significantly distinguish between the medium and highest quality. This suggests that some computational savings can be eventually gained with little perceptual loss in quality. Moreover, results from the group that experienced the real-world scenario and the control group that did not are not statistically significant indicating future studies of driving scenarios could be conducted entirely in virtual environments. Finally, findings suggest that participants give different judgements of quality when analysing visuals versus the overall experience.

The paper is divided as follows. An overview of physically-based rendering and related work is presented in Section II. The overall experimental methodology is given in Section III and the content stimuli generation in Section IV. Section V and Section VI present the results and discussion section respectively.

II. BACKGROUND AND RELATED WORK

This section presents the background on physically-based rendering and the related work that deals with comparing different quality settings for graphics.

A. Rendering

The process of simulating light transport in order to form images is known as rendering. This process takes an input consisting of the geometry of the virtual environment, materials which are applied to the geometry which specify the appearance of the surfaces, light sources to illuminate the scene, and a virtual camera on which to form the image. Two types of illumination are commonly considered; local illumination and global illumination. Local illumination consists of light reflecting once between the light source and the camera, and global illumination considers all the possible inter-reflections between objects in the scene before entering the camera. Outgoing light at each reflection event is commonly expressed as the solution to the Rendering Equation [6]:

$$L_o(x,\omega_o) = L_e(x,\omega_o) + \int_{\Omega} L_i(x,\omega_i) f_r((x,\omega_i,\omega_o)|N\cdot\omega_i|d\omega_i.$$
(1)

This states that the reflected radiance L_o from a point in the scene x in a direction ω_o is the sum of the emitted radiance L_e and the reflected radiance L_i from every visible point in the scene in the case of global illumination, or the light source in the case of local illumination. This is expressed as an integral over the hemisphere Ω aligned with the normal N at point x. This reflected radiance is weighted by the material properties at the point x, which is known as the Bidirectional Reflectance Distribution Function (BRDF) $f_r(x,\omega_i,\omega_o)$ and a cosine falloff to account for Lamberts Cosine Law (expressed as a dot product in Equation 1). The BRDF can be expressed as measured values encoded in either a 3D or 4D table [7], or as an analytic function. Typical examples include the diffuse surfaces (with a constant value of $\frac{\rho}{\pi}$), the Phong BRDF [8] and Microfacet models [9]. This equation can be extended over the sphere of directions which enables simulation of glass and

Fig. 1. The simulator used in the experiment. Note, in the experiment participants acted as passengers and the steering wheel column was removed.

scattering within the the surface. As the reflected light L_o in Equation 1 depends on the incident light L_i, which itself is the reflected light from another surface, this equation is recursive and has infinite dimensions.

Therefore, in order to compute images, Equation 1 has to be solved using numerical methods, most commonly Monte Carlo integration. This creates random samples from the integrand, averages them, and converges to the correct result as the number of samples tends to infinity. The error in the estimate appears as variance (noise) in the image. Many methods have been proposed to reduce this noise, see, for example, PBRT [10]. The Monte Carlo method used in this work is discussed further in Section IV-A.

Equation 1 can be split into a sum of two parts depending on the the source of the incident lighting L_i, corresponding to the local and global illumination components discussed above. This latter case simulates lighting effects such as light bleeding, caustics, and in the case of this work this is especially apparent as light entering a car through the windows and bouncing around inside the car. Both these cases can again be solved using Monte Carlo integration, and have the advantage that specialised approaches can be used which are more efficient at simulating direct (for example the work by Shirley and Wang [11]) and indirect lighting (for example the work by Veach [12]).

Two common simplifications to the approach of computing imagery are to use Whitted Ray Tracing [13] which calculates bounces between specular surfaces, but terminates paths and computes direct lighting when a ray hits a non-specular material, and ray casting which simply computes the light reflected off the nearest visible surface. Ray casting methods therefore do not compute reflection in mirrors or refraction through glass, which Whitted Ray Tracing does compute. Neither of these approaches compute full global illumination, but have a substantially lower computational cost.

B. Evaluations of Rendering at Different Qualities

While comparisons between differing methods are common in the imaging field, for example for High Dynamic Range (HDR) video compression [14] and image re-targeting [15], the research into evaluating participants' experiences of different rendering settings in VEs is not exhaustive. In the following the major related work in the area is outlined.

Fig. 2. A rendering of the virtual track used in the experiment.

Fig. 3. The real world track and car model used for the driving experience.

Zimmons and Panter [16] presented participants with the pit room experiment [17] across five different render settings, ranging from simple smooth shaded graphics to a radiosity solution (an alternative global illumination method that uses finite element methods [18]) , that reproduced indirect diffuse calculations, and found no significant differences across the rendering qualities, although all rendering qualities witnessed an increase in physiological measurements when entering the pit room.

Slater et al. [19] compared ray casting (see Section IV-A1) and ray tracing (see Section IV-A2) within VEs and found a significant effect of ray tracing versus the ray casting method. In a follow up experiment [20] the authors found different results when adding global illumination and no significant differences were found, suggesting that the main, earlier, difference was due to the dynamic shadows and reflections.

Slater et al. [21] allowed participants to modify the quality of certain settings and the illumination from plain shading to more complex lighting effects including dynamic lighting and found that under the condition of the feeling of reality the effect of illumination was considered important but was not as substantial from the alternative condition of direct presence.

Clearly, the above findings are inconclusive. In contrast with the above work, our highest quality settings are based around path tracing [6] which is an unbiased rendering method and offers a more accurate simulation of light transport. Furthermore, our work is centered around the concept of driving and how realistic the VEs are compared to real-world conditions.

III. METHOD

The experiment investigates whether a state-of-the-art global illumination method provides a significant improvement in

Fig. 4. The scale used in the experiments with textual recommendation.

the reported experience of participants in a driving simulator. Results can help define the fidelity required by future driving simulator systems. The experiment is based upon a real-world driving experience, see Figure 3, followed by a reproduced experience in a simulator. This section introduces the design of the experiment, the participant information and the procedure. Due to the complexity of the stimuli created, a separate section is dedicated to materials in order to delve further into the details.

A. Design

The goal of this experiment is to verify whether increasing the quality of the graphics rendering for driving simulators affects the overall perceived fidelity of the simulation. In order to achieve this goal a prototype driving simulator was built, see Figure 1. Participants were driven around in the simulator as passengers, and a group of the participants experienced the same track as passengers in a real-world reconstruction of the same drive and car. Participants were chosen as passengers, as opposed to drivers, due to the amount of computation time required to simulate real-world lighting accurately which currently prohibits real-time performance. The simulated experience consisted of simulated graphics via a head-mounted display, acoustics via headphones and haptics via a motion platform.

The participants were randomly divided into two groups accounted for by the between-participants independent variable *experience*. *experience* consisted of two conditions defining the two groups *real* for the participants that experienced the real-world drive before using the driving simulator and *virtual* for those participants that did not experience the real-world scenario but the virtual only. This was done to identify whether participating in a real-world experience representing the tested virtual environment had any effect on decisions in the virtual world.

Three rendering methods, corresponding to different stages of computation for simulating light transport introduced in Section II-A have been adopted for this experiment. This is represented by a within-participant independent variable *graphics* which consists of three possible conditions (*RC, WH, GI*) corresponding to ray casting (*RC*), Whitted ray tracing (*WH*) [13] and global illumination (*GI*) [6]. The amount by which these methods simulate the light transport will be described and shown in Section IV. The order of the first shown method from *graphics* was alternated such that they were balanced among the three variables and the watching of the video. The order of the rest of the presented variables was randomised.

A third, within-participants independent variable *modality* identified the ratings of either the participants' rating of the overall experience in a virtual environment (*VEexp*) or their

rating of the visuals alone (*Gexp*). The questions asked for *VEexp* and *Gexp* are outlined in Section III-C.

The design of the method is based around a rating design since it provides a degree of difference among the variables. The dependent variable was the value between 1 and 100 assigned by the user to the experience. The scale was displayed with verbal cues as shown in Figure 4.

There were two general hypotheses:

Hypothesis I: different graphics qualities will provide significant differences among the methods.

Hypothesis II: the overall experience judgment will differ from the overall graphics quality judgement.

Other data was also collected from the participants. This was collected after the first *graphics* variable was shown. These variables were the quality of the motion and acoustics with possible values between 1 and 100, again following the scale shown in Figure 4. The simulator sickness questionnaire (SSQ) [22], a motion sickness questionnaire that reports values between 0 (none) and 3 (very strong) on a number of simulator sickness characteristics, was also administered.

B. Participants

A total of 45 participants from an automotive manufacturer and a university took part in the experiment. They exhibited normal or corrected to normal vision. The *real* group consisted of 24 participants and the *virtual* group of 21. Participants exhibited an average age of 30.68 with a range from 20 to 66. Four participants did not report their age, as this was optional. One participant from the *real* group withdrew from the driving simulator due to motion sickness and this data was withdrawn from the results and analysis.

C. Procedure

Participants were randomly divided into a *real* and *graphics* groups. The *real* group were taken to the track in Gaydon via organised transport. After giving their informed consent, they were instructed to sit in the car as passengers and were driven around a chosen section of the track. Within a week of their participation, where timing was based on participant availability, they were asked to visit the virtual experience. The *virtual* group were only shown the virtual experience. The virtual part of the experiment consisted of participants being shown all three of the graphics scenarios and asked to mark the overall and graphics experience for each one right after that experience. The questions asked to the participants were: "On a scale of (1 ... 100) how close **was/were X** compared to the real world?" where "**was/were X**" was set to "were the visuals" for *Gexp*, "were the sounds" for acoustics, "was the motion" for motion, and "was the overall experience" for *VEexp*. Questions were asked in the order given above. Participants were also shown the scale in Figure 4 as guidance. Data for acoustics, motion and motion sickness (using SSQ) were collected only after the first experience. Participants were asked to sit in the motion platform and the head-mounted display and headphones were placed on them by the investigator. After experiencing all conditions they were asked whether they wanted to review their results and make any changes. Both sets of results were recorded.

IV. MATERIALS

This section introduces the details used for the driving experience, including the rendering methods used, geometry and track generation and the acoustic characteristics of the system.

A. Rendering Methods

This experiment compares animations created with three different rendering methods based on the theory discussed in Section II-A. These methods represent the quality levels which may be achieved with varying amounts of computation. Note all the rendering methods performed computation with the full track and car geometry. This consisted of 30 million unique triangles, and considering instancing, totalled approximately one billion triangles. No geometric simplification was performed in order for quality to only be affected by computational differences. For all methods every frame of the animation was rendered individually.

This section describes these methods, and the motivations for choosing each rendering quality. Each of these methods is based on ray casting, which is a procedure of determining visibility from a point (that is the first point the a scene from the start point of a ray, in the direction in which the ray is oriented), or if two points are mutually visible, as introduced in Section II-A. An overview of the three rendering methods is shown in Figure 5.

1) Low Quality - Ray Casting (RC): The low quality imagery used in this experiment consisted of a significant simplification to the rendering approach presented in Section II-A. This method is based around the visibility method used in ray tracing applications called Ray Casting (*RC*). The lighting in the scene consisted of a single directional light source, representing the sun. However, visibility tests for direct lighting were neglected, and the shading model was simplified. The resulting lighting at each visible surface was computed as follows:

$$L_o(x, \omega_o) = L_e(x, \omega_o) + \frac{L_A \rho_d}{\pi}$$
$$+ \ max \left(\frac{\rho_s(e+2)(R(N, \omega_o) \cdot \omega_s)^e}{2\pi}, 0 \right), \quad (2)$$

where ρ_d and ρ_s are diffuse and glossy reflectivity of the surface, e is an exponent to control how glossy a surface appears. L_A represents the ambient lighting at the surfaces not oriented in the direction of the sun ω_s. The BRDF used is a normalised Phong model [8]. Example frames of the animation are shown in Figure 6(a).

2) Medium Quality - Whitted Ray Tracing (WH): The imagery used for the medium quality animation was based on Whitted Ray Tracing (*WH*) [13]. In this approach, the sun was again approximated as a directional light source, and the previously described reflection model for surfaces was used. However, if the first surface visible to the camera is specular (glass or mirror), the ray tracing method spawns secondary rays which reflect or refract according to the material properties. This improves the previous approach by allowing light to reflect or refract on the windscreen, thereby adding veiling
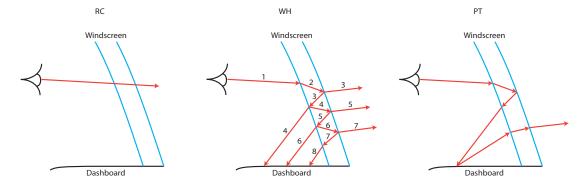
Fig. 5. Overview of the three rendering methods. The image on the left shows ray casting (*RC*) which ignores reflections on the windscreen, and only intersects the vehicle and scene geometry. The middle image shows Whitted Ray Tracing (*WH*) illustrating reflections and refractions through the glass of the windscreen. This image also illustrates why eight bounces were chosen, as this is able to capture veiling glare from multiple reflections in the glass. the right image shows path tracing which simulates the full range of lighting effects (*GI*), this example shows light reflecting and refracting through the glass, then reflecting off the dashboard back through the glass.

glare to the image. Additionally, visibility between the light source and the first non specular surface is computed in order to compute shadows. To compute shadows in the car, the ray casting method for medium quality frames ignores glass in the windows, otherwise the entire inside of the car would appear shadowed. Finally, this method uses eight rays per pixel distributed in a low discrepancy pattern to reduce aliasing in the image, and uses a maximum path depth of eight bounces in order to be able to capture veiling glare resulting from up to five total internal reflection events within the glass of the windscreen, see the middle image of Figure 5. Representative frames are shown in Figure 6(b).

3) High Quality - Global Illumination (GI): High quality rendering used Path Tracing [6], an unbiased method for solving the Rendering Equation (Equation 1). This method is termed Global Illumination (*GI*) in the following. This algorithm first samples a random point in a pixel, and then traces a ray from the virtual camera into the scene. Then the intersection with the first visible surface is computed and direct lighting is calculated. Next, the algorithm samples a probability density function associated with the material at the intersection to calculate indirect lighting. This process of computing the direct then indirect lighting is continued along a path until a stopping condition is met. Russian Roulette is used for this in order to ensure an unbiased result. At each intersection with the scene, the direct and indirect lighting is modulated by the BRDF and cosine term, and accumulated into the the amount of radiance which contributes to a pixel. This process is repeated many times per pixel in order to reduce noise from the Monte Carlo sampling, and is performed for all of the pixels and frames. $2,048$ samples per pixel were used to compute the images in this work, and followed by the application of a temporal filter to the results to remove any remaining noise. Two resultant frames are shown in Figure 6(c). These images show lighting which has entered the car through the glass in the windows, accurate veiling glare on the windscreen, and soft shadows from the sky.

B. Car Geometry and Materials

1) Geometry: The geometry for the vehicle was provided by Jaguar Land Rover (JLR), this is shown in Figure 7. The model comprises roughly 19.5M triangles.

2) Materials: Materials assigned to the car geometry were based on measured material samples that were provided by JLR.
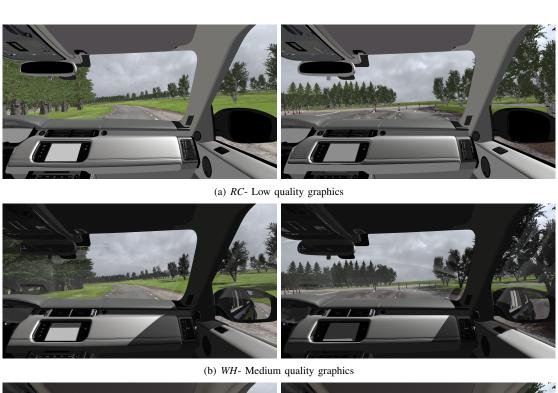
These samples were representative of the particular vehicle colour code of the vehicle used in the experiment. A planar sample of each material from the interior of the vehicle was measured as well as the visible components outside of the vehicle, for example the paint.

This measurement process was conducted using a gonioreflectometer [23], a device capable of measuring the BRDF of a particular material. This device captured a discrete approximation of the BRDF for a particular material. In order to use this measured BRDF data in the rendering system, the BRDF was fit to an analytical representation by following the procedure by Löw et al. [24]. The measured data was fit to a normalised Phong model for the Ray Casting (Section IV-A1) and Ray Tracing (Section IV-A2) methods, and was fit to a Microfacet model for use in the Global Illumination methods (Section IV-A3).

An editor for material assignment was created which tagged the geometry with the necessary metadata to be able to sample the correct virtual material.

C. Track Generation

The tesselated model of the track was generated through a process of re-topologising a captured point cloud into triangulated geometry. The point cloud scan comprised 88 individual scans with a resolution $40,000 \times 20,000$ points, see Figure 9. Neighbouring scans where controlled in such a fashion that reference targets (calibrated spheres) for aligning subsequent scans were visible and close enough that they would receive enough point information upon their surface, > 60 points, that they could be used as control points for limiting degrees of freedom between scans and stitching the scans together. This process was achieved using correspondence matching, shown in Figure 8.

(a) *RC*- Low quality graphics

(b) *WH*- Medium quality graphics

(c) *GI*- High quality graphics

Fig. 6. Examples of the presented visuals at different qualities.

Fig. 7. A view of the geometry provided by Jaguar Land Rover, with a default Lambert shader applied.

Fig. 8. Calibrated spheres showing their use in aligning the neighbouring scans by fitting scan points on their surfaces to one another, minimising total Euclidean distance.

The point cloud was homogenised to maintain point density across the scan, this process was achieved by culling points using statistical inference of their likelihood to be noisy and their local point density R, calculated through Equation 3. A point was deemed to be noisy if, by evaluating it's local neighbourhood, it could be adjudged to be an outlier to the point set by some variable radius. This happened if, for example, the laser scanner hit a dust particle and returned a floating point.

$$R = \frac{2\pi}{hd} \qquad (3)$$

where R is point density in metres, h is the horizontal resolution of the point cloud scan and d is the distance of the recorded point from the scanner position. For example,

Fig. 9. The aligned point cloud scan, comprising the set of 88 stitched individual scans.

given a scan resolution of $8,000 \times 4,000$ points, the point density at 10 metres from the scanner is 7.85 mm and 15.7 mm at 20 metres from the scanner.

Once the point cloud was processed as above, the information was a large set of points P. $\forall p \in P : \exists \{x, y, z, r, g, b\}$ where x, y, z is the 3D point location in Cartesian coordinates and r, g, b is the colour triplet of the point p. The procedure of converting this information from a collection of points p_n, $n = \{1 : 88 \times 40,000 \times 20,000\}$, was performed by projecting every point onto the ground plane along the vertical axis recording the colour and distance information into two separate textures. One was used for a displacement map and one to texture the displacement map, these are shown in Figure 10.

Fig. 10. Texture and displacement maps.

A height map used to create an output polygonised mesh would, in this instance, result in a few undesirable properties. Notably, high polygon count and a high amount of variance in the mesh around tree positions due to height differences in the point cloud scan in and around tree regions. Instead of using this process, the trees were filtered from the output mesh at the expense of some local accuracy by using a lofted spline approach to reconstruct the topology of the terrain. This consisted of a number of Bézier curves with equal number of control vertices. These were then lofted to create the profile of the terrain in Non-rational Uniform B-Spline format, which

was converted to a triangulated set of polygons. This had the effect of removing important but noisy details, such as signage, trees and foliage. These were re-injected into the final mesh as individual asset components.

D. Capture Information

1) Acoustics: Sound was recorded during a test drive for the real condition. The sound recording was performed using a TetraMic [25], which is an Ambisonic soundfield microphone with four capsules arranged in a tetrahedral fashion that record in A format. This data can be readily converted into B format [26] (W, X, Y, Z) and then interpreted into any single point configuration of microphones. The TetraMic was placed with the front facing vector of the microphone similar to that of a conventional passenger.

The choice was made to use headphones and output to binaural using a Head Related Transfer Function (HRTF) as the user would be wearing a Head Mounted Display (HMD) which provided head tracking information. The head tracking information can be used to decode the B format file to a binaural stream via the HRTF to produce the correct binaural response given the particular view direction in the sound field. The KEMAR HRTF data set was used for this process [27].

2) Reconstructing Drive Path: The capture of the sound of the real drive happened simultaneously with a Global Positioning System (GPS) data capture process, which were both later used in the virtual drive reconstruction. A base station at the track allowed capture of real-time six degree-of-freedom (DOF) motion data for the vehicle, and also the world position in terms of latitude and longitude. This information, due to the geodetic nature of the point cloud scan, was used to align the real recorded drive, with the reconstructed virtual drive path. The precision of this process is limited by the precision of the base station, which is quoted as $\pm$2cm. Conversion of the recorded GPS latitude and longitudinal data to Universal Transverse Mercator (UTM) coordinates was performed to create an accurate camera path for the virtual drive imagery. The precision of this conversion process is accurate to within 1mm [28]. The entirety of the track resided within UTM zone 30U. The GPS signal for the real drive was recorded at 100Hz, the virtual drive imagery was to be displayed at 24Hz; interpolation occurred to reduce the 100Hz GPS signal and resample to 24Hz when calculating the position of the vehicle in accordance with the resultant frame rate of the imagery for the virtual drive. These positions were used to derive the temporal camera path by projecting the 24Hz UTM positions onto the geodetic track data. This process supplemented the Easting and Northing coordinates of UTM (in metres) with the absolute height (in metres) from the measured track data.

V. RESULTS

Results are presented with captured and normalised data. Normalised data is normalised from the captured values between 1 to 100 to account for individual differences between the participants. The normalisation was conducted as:
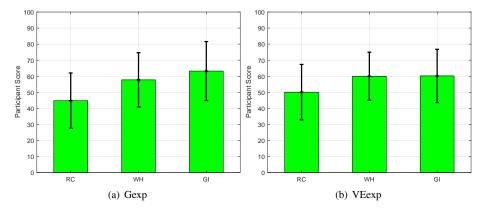
$$rn_{pij} = \frac{r_{pij} - \mu_p}{\sigma_p} \qquad (4)$$

(a) Gexp   (b) VEexp

Fig. 11. Mean and standard deviation for *graphics*.

TABLE I
MEANS FOR ALL CASES ANALYSED.

	Captured Data (/100)						Normalisation					
	Visuals (*Gexp*)			Overall (*VEexp*)			Visuals (*Gexp*)			Overall(*VEexp*)		
	RC	WH	GI	RC	WH	GI	RC	WH	GI	RC	WH	GI
real	44.57	53.78	63.09	48.91	55.52	57.47	-0.68	-0.12	0.46	-0.33	0.04	0.16
virtual	45.00	62.05	63.38	51.33	64.90	63.14	-1.11	0.07	0.27	-0.57	0.30	0.30
Overall	44.77	57.73	63.22	50.07	60.00	60.18	-0.88	-0.03	0.37	-0.45	0.16	0.23

where r_{pij} was the rating for graphics method i (where i is *RC, WH, GI*) for participant p and for modality j (where j is either *VEexp* or *Gexp*), μ_p and σ_p are the mean rating and standard deviation for participant p for all r_{pij} and rn_{pij} is the resulting normalised variable. The normalised results are used for the inductive statistics below.

Means for all the cases are shown in Table I. This includes the data that was recorded and the normalised data via the use of Equation 4. Results of the raw data, see Figure 11, show that the methods elicited results that for visuals (*Gexp*), on average, would be on the scale of similar to very similar (50 - 75) for *WH* and *GI* and on the scale of slightly similar to similar (25 - 50) for *RC* (see Figure 4). For overall (*VEexp*) the difference was smaller and all graphics conditions range from similar to very similar in the response scale. Results show the ordering of preference is *GI, WH, RC*. A statistical analysis of whether significant differences arise in these results is given below.

A. Overall Analysis

The overall results were analysed using a 2 (*experience*) × 3 (*graphics*) × 2 (*modality*) repeated measures factorial ANOVA. The F-test, the standard test in ANOVA analyses, tests whether the means of several groups differ from each other. More formally it is defined as:

$$\frac{\sum_{k=1}^{K} n_k(\mu_k - \mu)/(K-1)}{\sum_{k=1}^{K}\sum_{l=1}^{n_k}(x_{kl}-\mu_k)/(N-K)},$$

where, N is the total number of observations, μ the mean across all observations, K the number of groups, n_k the number of observations in group K, μ_k the mean of group K, and x_{kl} the lth observation of group K. F is reported as F($K-1$, $N-K$) corresponding to the degrees of freedom under the null hypothesis.

The effect of *experience* was found to not be significant F(1,42) = 0.216, p = 0.66. This indicates that there was no effect of whether participants attended the real-world scenario first or did not attend the real-world experience at all.

The effect of *graphics* was found to be significant F(2,84) = 25.85, p < 0.01. The means of the *graphics* were: *GI* having the overall highest rating (μ_{GI} = 0.297), followed by *WH* (μ_{WH} = 0.073) and finally *RC* (μ_{RC} = -0.673). Pairwise comparisons, with Bonferroni corrections, among the three methods shows significant differences between *RC* and both *WH* and *GI* but no significant difference between *WH* and *GI*. This indicates that while Hypothesis I is accepted it only partially holds; there is an overall difference among the qualities and the higher quality graphics are preferred in terms of ordering but there is no significant difference between *WH* and *GI*.

The effect of *modality* was not found to be significant F(1, 42) = 3.46, p = 0.07 . However, there was a significant effect on the interaction *modality* × *graphics*, F(2, 84) = 6.55, p < 0.01, suggesting that, as expected, judging visual versus the overall experience produced different results. Figure 12 shows this interaction; the overall values appear to move closer to each other for the *modality* variable of *VEexp*.

The significant interaction *modality* × *graphics* requires further analysis of the results [29] at each level of *modality* for *Gexp* and *VEexp* individually in order to better understand the relationship of *graphics* across *modality*. This analysis is carried out in separate ANOVAs below.

B. Graphics experience only results

For *Gexp* the results were analysed using a 2 (*experience*) × 3 (*graphics*) repeated measures factorial ANOVA.

As with the overall effect, the effect of *experience* was found to not be significant F(1,42) = 1.07, p = 0.31.
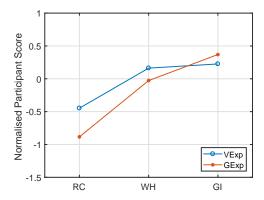
Fig. 12. *Modality* × *graphics* showing the significant interaction between the two for normalised scores. Means of normalised data are shown.

TABLE II
MEANS FOR ACOUSTICS AND MOTION.

	RC	WH	GI
Acoustics	51.92	46.40	53.09
Motion	56.92	52.20	46.36

The effect of *graphics* was found to be significant F(2,84) = 31.89, p < 0.01. Pairwise comparisons, with Bonferroni corrections, among the three methods were significant. The means for *GI* were rated highest (μ_{GI} = 0.365), followed by *WH* (μ_{WH} = -0.027) and finally *RC* (μ_{RC} = -0.884).

Furthermore, for this analysis, and the subsequent analysis of *VEexp* below, Kendall's co-efficient of concordance (W) was also analysed. Kendall's W evaluates how close the participants' order of the results were to each other whereby a value of 1 indicating identical ordering and 0 complete disagreement. Kendall's W was 0.34, significant at p < 0.01.

C. Virtual reality experience only results

As with *Gexp*, *VEexp* the results were analysed using a 2 (*experience*) × 3 (*graphics*) repeated measures factorial ANOVA.

As with the overall effect, the effect of *experience* was found to not be significant F(1,42) = 0.16, p = 0.69.

The effect of *graphics* was found to be significant F(2,84) = 10.51, p < 0.01. Pairwise comparisons, with Bonferroni corrections, among the three methods were significant for *RC* and the other two methods but not for *GI* and *WH*, following the overall results. The means were closer for *VEexp* than *Gexp* but followed the same order; *GI* rated highest (μ_{GI} = 0.229), followed by *WH* (μ_{WH} = 0.162) and finally *RC* (μ_{RC} = -0.447).

Kendall's W for was 0.15, still significant at p < 0.01 but lower than the Kendall co-efficient for *Gexp* indicating more indecision among the participants choices.

D. Further Results

Means for acoustics and motion are shown in Table II. The quality of acoustics and motion were analysed via a one-way ANOVA across *graphics*. For acoustics no significance was found F(2,32) = 0.74, p = 0.49. Similarly, for motion no significance was found F(2,32) = 0.64, p = 0.54.

TABLE III
MEANS FOR SSQ CHARACTERISTICS NAUSEA (N), OCULOMOTOR (O), DISORIENTATION (D) AND TOTAL SCORE (TS) (KENNEDY ET AL. [22]) FOR ALL *graphics* CONDITIONS.

	RC	WH	GI
N	19.08	12.14	20.82
O	20.41	17.92	22.05
D	33.19	27.84	21.51
TS	271.83	216.54	240.77

For simulator sickness, the results from SSQ, where aggregated into results for Nausea (N), Oculomotor (O), Disorientation (D) and Total Score (TS) following Kennedy et al. [22]. Table III shows the means for N, O, D and TS. Furthermore, all four parameters where analysed independently via the non-parametric Kruskal-Wallis test, due to the use of ordinal values in the data capture. For all parameters no significance across *graphics* was found. For N, H(2) = 1.31, p = 0.52; for O, H(2) = 1.74, p = 0.42; for D, H(2) = 0.91, p = 0.63; and for TS, H(2) = 1.36, p = 0.51.

Finally, a test is carried out to verify whether the length of time between the real world experience and the virtual world experience had any effect on the resultant scores for the *real* group. The length of days from the real world experience to the virtual world experience was correlated with the average score given by the participants and found to be insignificant, r = -0.033, p = 0.884.

VI. DISCUSSION

Results show that Hypothesis I is accepted. However, while there is a significant difference among the three methods, across all the results, this difference is only noted between the *RC* and the rest of the results. The distinction across the two modalities indicates that this is mostly due to the participants' judging of the overall experience in which the differences between the *WH* and *GI* were insignificant. The differences across all the qualities were observed when judging only graphics and this demonstrates an observable difference between the methods. Yet, in nearly any application participants are engaged in an activity, and the ability to discriminate between the quality decreases and quality changes may not be observable. Interestingly, this compares strongly with results from other work. In particular the combined experiments of Slater et al. [19] and Yu et al. [20] who also found that some form of advanced graphics based on reflections and shadows (similar to *WH*) provided an enhanced experience compared to the basic rendering, however no differences were found with the more advanced full global illumination results.

Hypothesis II was not accepted, no significant difference was noted between scores for overall and visual fidelity in particular. However, importantly, the interaction between *graphics* × *modality* was significant. This led to identifying that a difference in graphics is noted, as discussed above, when evaluating graphics only but not when evaluating the overall experience. As mentioned above, nearly all interaction involves participants experiencing complex environments and results indicate that the lack of a notable difference means

that, once such quality levels can be achieved, computational times can be saved with little perceivable loss of quality.

It is important to note that the *RC* method corresponds to many of the current driving simulator illumination settings. *RC* has performed significantly worse under all scenarios, so there are still improvements to be made in order to deliver higher fidelity experiences and the investment in further computation would be deemed beneficial. Moreover, *RC* presented here may be judged to be higher quality due to the complexity of geometry used, two or more orders of magnitude more complex than most driving simulators. Descriptive results for *RC* were judged to be similar to the real world on average.

The comparison between the real world and virtual experience result was not considered significant. While no conclusions could be drawn from the lack of significance, nearly all participants have experience as drivers and/or passengers, hence it is unsurprising that no differences are found. In future, the use of a real-world experience is probably not required, especially considering the complex arrangements required to conduct such real-world studies.

A. Limitations and Future Work

The subdivision of graphics quality as used in this work is based on three fixed states. The quality of the visual simulation can be more complex and fit a more continuous scale and this work could be considered to have been limited to three fixed points on such a scale. The chosen settings do, however, correspond broadly to general settings used, particularly the low quality is the same level of shading found in many simulators, and the highest quality path tracing corresponds to state-of-the-art methods as witnessed in lighting simulations and high-end entertainment [5]. These settings were originally chosen because they broadly reflect such general settings. Future work will look into more subtle control of quality changes expanding on the general approach presented by Slater et al. [21] and analyse rendering parameters such as lighting, shadows, direct vs. indirect lighting, scattering etc. Furthermore, the current approach employed a single BRDF model based on measuring the materials, more complex BRDF models could be employed and analysed. An approach for analysing across different parameters similar to that adopted by Doukakis et al. [30] for analysing the contribution of audio-visual cross-modal contributions to virtual environments or Debattista et al. [31] for analysing the contributions of frame rate against resolution could be adopted across further rendering settings. Finally, the use of on-screen avatars is known to have an effect on participants' immersion in virtual environments, and avatars will be considered in future work.

The quality of the visuals and overall experience are not just based on a function of the illumination. Firstly, visual fidelity can depend on a number of other factors such as geometry, textures etc. In this work the quality of the geometry and textures were fixed such that only lighting changes were investigated. Yet, the quality of all parameters was much higher than is typical for driving simulators so, potentially, the lower and medium quality results could have been automatically boosted in the presented results. The difference in results indicate that there are still differences in quality of illumination but these advantages must be considered for the lower qualities when analysing this work. Secondly, the perception from stimuli across the senses combine to produce a cross-modal experience where certain modalities are attended to more than others [32]. This work looked at fixing the input from the other senses and thus controlling for them, using state-of-the-art acoustics and a practical motion platform for haptic feedback, as such no significant differences were noted for these stimuli across the different qualities. However, research in the area of cross modality and how that affects rendering of VEs is growing [33] and aspects of cross-modal interaction will be included in future work.

REFERENCES

[1] D. A. Bowman and R. P. McMahan, "Virtual reality: how much immersion is enough?" *Computer*, vol. 40, no. 7, 2007.
[2] C. Ariën, E. M. Jongen, K. Brijs, T. Brijs, S. Daniels, and G. Wets, "A simulator study on the impact of traffic calming measures in urban areas on driving behavior and workload," *Accident Analysis & Prevention*, vol. 61, pp. 43–53, 2013.
[3] D. Hallvig, A. Anund, C. Fors, G. Kecklund, J. G. Karlsson, M. Wahde, and T. Åkerstedt, "Sleepy driving on the real road and in the simulatora comparison," *Accident Analysis & Prevention*, vol. 50, pp. 44–50, 2013.
[4] V. Faure, R. Lobjois, and N. Benguigui, "The effects of driving environment complexity and dual tasking on drivers mental workload and eye blink behavior," *Transportation research part F: traffic psychology and behaviour*, vol. 40, pp. 78–90, 2016.
[5] C. Eisenacher, G. Nichols, A. Selle, and B. Burley, "Sorted deferred shading for production path tracing," in *Computer Graphics Forum*, vol. 32, no. 4. Wiley Online Library, 2013, pp. 125–132.
[6] J. T. Kajiya, "The rendering equation," in *ACM Siggraph Computer Graphics*, vol. 20, no. 4. ACM, 1986, pp. 143–150.
[7] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," in *ACM SIGGRAPH 2003 Papers*, ser. SIGGRAPH '03. New York, NY, USA: ACM, 2003, pp. 759–769. [Online]. Available: http://doi.acm.org/10.1145/1201775.882343
[8] E. P. Lafortune and Y. D. Willems, "Using the modified phong reflectance model for physically based rendering," 1994.
[9] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, "Microfacet models for refraction through rough surfaces," in *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, 2007, pp. 195–206.
[10] M. Pharr and G. Humphreys, *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2004.
[11] P. Shirley and C. Wang, "Direct lighting calculation by monte carlo integration," in *Photorealistic Rendering in Computer Graphics*. Springer, 1994, pp. 52–59.
[12] E. Veach, "Robust monte carlo methods for light transport simulation," Ph.D. dissertation, Stanford University, 1997.
[13] T. Whitted, "An improved illumination model for shaded display," *Commun. ACM*, vol. 23, no. 6, pp. 343–349, Jun. 1980.
[14] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers, "Objective and subjective evaluation of high dynamic range video compression," *Signal Processing: Image Communication*, vol. 47, pp. 426–437, 2016.
[15] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in *ACM transactions on graphics (TOG)*, vol. 29, no. 6. ACM, 2010, p. 160.
[16] P. Zimmons and A. Panter, "The influence of rendering quality on presence and task performance in a virtual environment," in *Virtual Reality, 2003. Proceedings. IEEE*. IEEE, 2003, pp. 293–294.

[17] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr, "Walking¿ walking-in-place¿ flying, in virtual environments," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 359–364.

[18] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces," in *ACM SIGGRAPH Computer Graphics*, vol. 18, no. 3. ACM, 1984, pp. 213–222.

[19] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment," *IEEE computer graphics and applications*, vol. 29, no. 3, 2009.

[20] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual realism enhances realistic response in an immersive virtual environment-part 2," *IEEE computer graphics and applications*, vol. 32, no. 6, pp. 36–45, 2012.

[21] M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 4, p. 92, 2010.

[22] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[23] S. C. Foo, "A gonioflectometer for measuring the birdirectional reflectance of material for use in illumination computation," Ph.D. dissertation, Cornell University, August 1997.

[24] J. Löw, J. Kronander, A. Ynnerman, and J. Unger, "Brdf models for accurate and efficient rendering of glossy surfaces," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 1, p. 9, 2012.

[25] D. Hemingson and M. Sarisky, "A practical comparison of three tetrahedral ambisonic microphones," in *Audio Engineering Society Convention 126*, May 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14872

[26] A. Farina. (2017) A-format to b-format conversion, Last Accessed: 27 Jan 2017. [Online]. Available: http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm

[27] B. Gardner and K. Martin, "Hrtf measurements of a kemar dummy-head microphone," MIT Media Lab Perceptual Computing, Tech. Rep., 1994.

[28] C. F. F. Karney, "Transverse mercator with an accuracy of a few nanometers," *Journal of Geodesy*, vol. 85, no. 8, pp. 475–485, Aug 2011. [Online]. Available: https://doi.org/10.1007/s00190-011-0445-3

[29] D. C. Howell, *Statistical methods for psychology*. Cengage Learning, 2012.

[30] E. Doukakis, K. Debattista, C. Harvey, T. Bashford-Rogers, and A. Chalmers, "Audiovisual resource allocation for bimodal virtual environments," *Computer Graphics Forum*, pp. n/a–n/a. [Online]. Available: http://dx.doi.org/10.1111/cgf.13258

[31] K. Debattista, K. Bugeja, S. Spina, T. Bashford-Rogers, and V. Hulusic, "Frame rate vs resolution: A subjective evaluation of spatiotemporal perceived quality under varying computational budgets," *Computer Graphics Forum*, pp. n/a–n/a. [Online]. Available: http://dx.doi.org/10.1111/cgf.13302

[32] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers, "Acoustic rendering and auditory–visual cross-modal perception and interaction," in *Computer Graphics Forum*, vol. 31, no. 1. Wiley Online Library, 2012, pp. 102–131.

[33] C. Harvey, K. Debattista, T. Bashford-Rogers, and A. Chalmers, "Multimodal perception for selective rendering," in *Computer Graphics Forum*. Wiley Online Library, 2016.

Thomas Bashford-Rogers is a Senior Lecturer at the University of the West of England, and has a degree in Computer Science from the University of Bristol and a doctorate in Computer Graphics from the University of Warwick. His research interests include global illumination, ray tracing and Monte Carlo methods. Contact him at Tom.Bashford-Rogers@uwe.ac.uk.
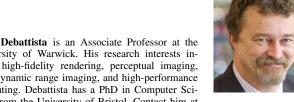

Carlo Harvey Carlo Harvey is a Senior Lecturer at Birmingham City University and a member of the research team at the Digital Media Technology Laboratory. He has a degree in Computer Science from the University of Bristol and a PhD in Engineering from the University of Warwick. His research interests include acoustics, perceptual rendering and cross modal interactions. Contact him at carlo.harvey@bcu.ac.uk.

Brian Waterfield is Virtual Reality and High-end Visualisation Technical Lead in the Research department at Jaguar Land Rover, after creating and installing the Virtual Innovation Centre at their design and engineering facility at Gaydon. He has moved onto the front end of the creation process, inventing and innovating the future virtual product and process enabling technologies. He is creating a new state of the art centre as part of the new National Automotive Innovation Centre, where he will continue to engage with universities and cross sector industries to drive his passion to one day immerse all the senses, to a level that creates a perceptual equivalent experience of the real world. He also co-founded ImmerseUK, which brings together industry, researchers and research organisations, the public sector, entrepreneurs, innovators and end users to support the UK in becoming the global leader in applications of immersive technologies: high-end visualisation, virtual, mixed, and augmented reality, haptics and other sensory interfaces with data. Contact him at bwaterf1@jaguarlandrover.com.

Kurt Debattista is an Associate Professor at the University of Warwick. His research interests include high-fidelity rendering, perceptual imaging, high dynamic range imaging, and high-performance computing. Debattista has a PhD in Computer Science from the University of Bristol. Contact him at K.Debattista@warwick.ac.uk.

Alan Chalmers is a Professor of Visualisation at WMG, University of Warwick, UK and a former Royal Society Industrial Fellow. He has a PhD from University of Bristol, 1991. He is Honorary President of Afrigraph and a former Vice President of ACM SIGGRAPH. Chalmers has published over 245 papers in journals and international conferences on HDR, high-fidelity virtual environments, multi-sensory perception, parallel processing and virtual archaeology and successfully supervised 43 PhD students. In addition, Chalmers is a UK representative on IST/37 considering standards within MPEG. Contact him at A.Chalmers@warwick.ac.uk