

# Fidelity and Plausibility of Bimanual Interaction in Mixed Reality

Gregory Hough, Ian Williams, and Cham Athwal

**Abstract**—When human actors interact with virtual objects the result is often not convincing to a third party viewer, due to incongruities between the actor and object positions. In this study we aim to quantify the magnitude and impact of the errors that occur in a bimanual interaction, that is when an actor attempts to move a virtual object by holding it between both hands. A three stage framework is presented which firstly captures the magnitude of these interaction errors, then quantifies their effect on the relevant third party audience, and thirdly assesses methods to mitigate the impact of the errors. Findings from this work show that the degree of error was dependent on the size of the virtual object and also on the axis of the hand placement with respect to the axis of the interactive motion. In addition, actor hand placement outside and away from the object surface was found to affect the visual plausibility considerably more than when the actor's hands were within the object boundaries. Finally, a method for automatic adaptation of the object size to match the distance between the actor's hands gave a significant improvement in the viewers' assessment of the scene plausibility.

**Index Terms**—Interactive virtual studios, human performance measurement, interaction framework, mixed reality, interactive virtual environment, interaction error, user study

---

## 1 INTRODUCTION

MIXED Reality (MR) environments are becoming increasingly commonplace in modern day media. Although most mixed reality applications simply overlay virtual objects onto real scenes, there is also a growing trend towards allowing interaction between humans and virtual objects, for example HandyAR [1]. In this paper, we consider the fidelity of direct interaction, namely where users employ both hands (bimanual) to physically move a virtual object in a mixed reality environment.

In particular, we focus on an MR environment designed for creating live television programmes within a Virtual Studio, where the real actor appears to exist inside a virtual set. Our previous work reported on constructing an Interactive Virtual Studio (IVS) to allow realistic interaction between the actor and the virtual objects, based on three main strands of research: ubiquitous visual feedback [2], automated occlusion handling [3] and interaction with virtual objects. In that work, case studies assessing the potential for interaction within a virtual TV studio identified that bimanual interaction was a frequent modality. That is where actors move, adjust and relocate a virtual object in real time by holding it with both of their hands. Videos of this system in operation are found at [4]. In these case study investigations, it was noted how actors frequently failed to position their hands correctly on the surface of the virtual objects, both at the start and throughout the interaction. Our further studies [5], [6] found, that as

the actors have no physical boundaries to aid their hand placement, they were likely to place them either outside, away from the surface of the object (we term this “overestimation” of the object size) or inside the object, therefore intersecting it (termed “underestimation”). To a third party viewer, for example a screen or TV viewer, these effects can make the interaction appear implausible [6]. A similar issue was documented by Minoh et al. [7] who recognised that misalignment between the actor and virtual object could negatively impact the plausibility of an interaction, from the perspective of a third person viewer. However, in that study the inconsistencies were primarily due to device errors and not the actor's motion.

The first stage (Section 3) of this paper analyses the motion of actors completing a series of interaction tasks under controlled conditions. From this, a profile of the likely errors is constructed and their causes analysed. The second stage (Section 4) analyses third party viewer perception of the errors. A series of videos depicting the type of errors found in Section 3 are presented to a group of viewers for rating. From these results the effect of the type of errors is quantified by a viewing audience. Combining the two sets of results allows an analysis of perceptually acceptable interactions to guide the design of suitable interaction scenarios.

Section 5 involves the assessment of two techniques that aim to mitigate the impact of the identified errors: (a) adapting the size of the interactive object based on any measured interaction errors and (b) adjusting the background colour. Section 6 concludes by presenting a list of recommendations which can aid the development of mixed reality environments offering improved bimanual interaction.

## 2 BACKGROUND

### 2.1 Measuring Performance in HCI

Many methods and performance metrics exist for assessing usability in interactive virtual environments (VEs), yet the

---

DMT Lab, Birmingham City University,

Birmingham B4 7XG, West Midlands. E-mail: [Gregory.hough2@mail.bcu.ac.uk](mailto:Gregory.hough2@mail.bcu.ac.uk), [Ian.Williams,Cham.Athwal@bcu.ac.uk](mailto:Ian.Williams,Cham.Athwal@bcu.ac.uk).

specific challenge of quantifying the motion of actors in MR/IVS environments has not yet been conducted. When investigating Human Computer Interaction (HCI) in VEs, it is typical to assess the usability of a system with performance metrics such as task completion time and task failure/error rate, as these are robust and portable descriptors. Meanwhile, within this work we require the design of performance metrics that are pertinent to a third person viewer's experience and allow analysis for improvement of actor performance.

Fitts' law was one of the earliest measurements of human motion in relation to HCI [8]. It describes the amount of time a person takes to point to or touch an object based on a function of distance, object width, and task difficulty. However, as Fitts' law is used to describe the time taken to complete a specific pointing task, it is not suited to predict hand placement accuracy to the surface—although the concept of object size is a feature common to this work.

Another example is Bowman et al. [9] who were able to compare user performance with different virtual reality interfaces using task completion time as a key performance metric. This is a frequent metric within VE interface assessment owing to the fact that most VE interfaces are concerned with ease of use, so the usability metrics tend to focus on how consistently and quickly the user is able to perform tasks.

Less frequently, distance from an ideal, such as a target location, is used as a performance metric. An example of this is with Teather and Stuerzlinger [10] who used distance error from a target location (alongside task completion time) to assess the accuracy of users in completing 3D tasks with a range of 3D and 2D interface devices.

Heuristic evaluation has also been shown to be effective in assessing performance of users in interactive VEs. An example of this is presented by Sutcliffe and Gault [11], who identified impacting conditions on user performance in VEs. This was achieved by drawing experience based qualitative conclusions from the performance of a series of common tasks.

Guiard's model of bimanual skill [12] describes the roles the hands have in asymmetric bimanual tasks. The theory states that the role of the non-preferred hand is to lead the preferred hand when performing tasks, whereas the preferred hand produces fine movement. The bimanual tasks presented in this body of work are symmetric with respect to both hands, meaning Guiard's model is not applicable in the context of this study.

## 2.2 Aiding Fidelity of Actor Motion in the Virtual Studio

Two key studies for supporting actor motion in the virtual studio are the work of Kim et al. [13] and Woldecke et al. [14], [15]. These analyse the effectiveness of using haptic devices to aid actors in locating the surfaces of virtual objects. Kim et al. [13] used vibrotactile haptic devices placed in the palm of the actor that vibrated when a collision with a virtual object was detected. While these helped improve the performance of the actor, they were still liable to incorrectly estimate the surface of the virtual object due to a slow response time from the devices. Woldecke et al. also developed a vibrotactile system, this time in the form

of belts placed around the waist [14] or arm [15] that provide pulsed haptic feedback to the actor. This method was effective in notifying an actor of the location of a nearby virtual object, however when this system was used for guiding the actor's arms it was found to be insufficient for arm coordination when used alone and supplementing it with visual feedback was recommended.

In the context of manual identification of object locations in a virtual studio, Woldecke et al. recognised that an actor would struggle to find discrete locations such as the surface of a static virtual object. This corresponds to the study presented in our paper, where the actor also attempts to place their hands near the surface of a virtual object. As such, the framework presented in our study can be seen as an expansion on Woldecke et al.'s work into the bimanual domain, which carries with it an extended range of potential actor error conditions.

## 2.3 Measuring Scene Plausibility

In our proposed scenario, the belief that the actor is making contact with the sides of the virtual object as they would with a real one is an element of the plausibility of the scene. Previous studies into analysing plausibility of MR scenes have focused on assessing the discernibility of virtual objects as opposed to real objects. Scene manipulation to reduce photorealism has been shown to be an effective method in reducing the discernibility of objects as being non-real in an augmented reality environment. Non-Photorealistic Rendering (NPR) (e.g., Haller [16] and Hertzman [17]) is a method of applying filters that produce "stylised" non-photorealistic effects in an attempt to normalise the real and the virtual objects in a scene. These filters typically place a thick border around detected edges and homogenise the textures in the scene, creating a painted effect. Conventionally this has been tested by presenting participants with real and virtual objects in an augmented reality scene and asking them to identify which ones are virtual. Fischer et al. [18] tested the effectiveness of NPR using 60 video sequences displayed on a monitor to 18 participants who were required to identify whether the object in the scene was real or virtual. 15 real objects and 15 virtual counterparts were shown using two render modes, conventional AR and stylized. Participants were able to discern between the objects in 94 percent of conventional cases presented, but only in 69 percent of cases with the stylised mode, where lower discernibility indicates that virtual and real objects appear more similar. Steptoe et al. [19] also examined real time NPR, here using an augmented reality headset, as opposed to the desktop based system of Fischer et al. [18]. 30 participants had to identify which of 10 objects placed in an area in front of them (five real, five virtual) were real using three modes of rendering, conventional, stylized and virtualized, using 10 subjects for each mode. Virtualised mode presented the outlines of the virtual and real objects only. Stylised (56 percent accuracy) proved to be more successful than conventional augmented reality rendering (73 percent), with virtualised outperforming both (38 percent).

Whilst the above studies have relied on binary identification of objects in mixed reality scenes, for traditional broadcast video there are standard methods, notably the International Telecommunications Union (ITU-R) BT.500

TABLE 1  
Motion Analysis Experiment Conditions

Condition	Definition	Levels
Size	The length of the virtual object in the axis of hand placement.	18.2 cm, 36.4 cm, 54.5 cm
Target speed	The speed at which the actor aims to complete the interaction	0.17 m/s, 0.26 m/s, 0.34 m/s
Axis of motion	The axis along which the interaction takes place	Horizontal (H), Vertical (V)
Axis of hand placement	The sides of the object the hands are placed on	Left and Right (LR), Top and Bottom (TB)
Direction	The direction of the interaction	To the Left, To the Right <i>or</i> Upwards, Downwards

[20], for analysing the end viewers' perceived quality of scenes. Typically used for assessing the quality of video codecs, the BT.500 methodology has been applied to measure the perception of spatial artefacts (e.g., blockiness or posterizing (Winkler et al. [21], Wang et al. [22], Wu et al. [23])), or temporal artefacts (e.g., dropped frames, such as Pastrana-Vidal et al. [24]).

Here it is important to demonstrate the transferability of BT.500 methodology as measuring fidelity between an actor's hands and the surfaces of a virtual object would be considered an atypical and novel application of the BT.500 recommendations. Here we introduce two examples that reflect its flexibility.

Deshpande [25] presented a method for assessing the perception of mismatched synchronization between the monitors in tiled video displays, a form of display where multiple monitors are placed together and show a single video image imitating a larger display. Using the BT.500 recommendations as a base for their methodology, Deshpande was able to successfully analyse the effect that mismatched synchronization between the tiles of the display had on observers. IJsselsteijn et al. [26] applied the BT.500 methods for evaluating the issue of "presence" when viewing stereoscopic displays. In this study 12 observers viewed three videos, continuously rating their sense of presence during the scene. The authors were able to identify the impact that certain visual artefacts associated with stereoscopy have on presence (e.g., inconsistent depth and occlusion cues) and when the observers' sense of presence was reduced.

### 3 ACTOR MOTION ANALYSIS

This section presents the results of a human motion study where a group of actors complete a series of interaction tasks in a simulated virtual studio environment. This study was conducted to quantify the extent of errors in hand placement when an actor attempts to move a virtual object by holding it on the sides.

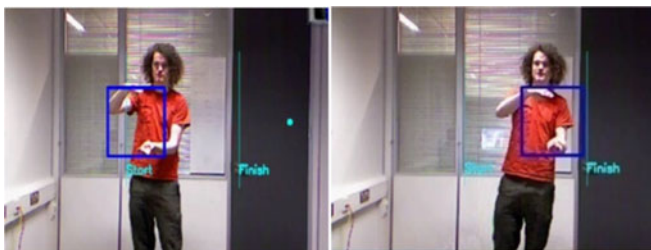


Fig. 1. Feedback presented to the subject during the interactive tasks. The markers indicate the start and end position of the task.

The interaction tasks were chosen based on a series of case-study scenarios that were part of the experimentation reported in our study of occlusion in virtual studios [3]. The most common interaction types required within these scenarios were bimanual interactions, where an actor is required to place their hands on the side of an object and move it in a specified manner. It was noted that actors frequently placed their hands incorrectly with respect to the object surface, and that during the interaction the hand placement was seen to vary. These errors in hand placement are likely to be detrimental to the quality of the end broadcast video.

In the following experiment, we test the null hypotheses that "the actor performance in maintaining their hands correctly at the surface of the virtual object is independent of: a) object size, b) target interaction speed and c) hand position."

Within this study, linear translation of an object pertaining to the conditions of Table 1 are evaluated. This allows us to provide definitive results within the scope of this set of common conditions, and a basis for future work to investigate other conditions such as more complex translations and rotations.

#### 3.1 Procedure

The equipment comprises of a 640\*480 pixel video camera, a Microsoft Kinect 1 motion capture device and a feedback monitor that is connected to a computer capable of layering virtual objects in front of the video feed. Fig. 1 presents an example of the feedback provided to the actor during the experiment. Shown in this figure are the start and finish locations of the intended interaction and the interactive object, which is locked to the midpoint of the hands. Because we are considering the creation of interactions that are to be viewed by a third party, it is not necessary for the interactive object to be fully rendered when presented for actor feedback. In this case we employ a 2D blue rectangle for two reasons. Firstly simple (2D) boundaries for objects provides good visual saliency to aid hand placement fidelity. Secondly feedback can be provided to the actor at a low latency (<40 ms), which otherwise might not be achievable if complex 3D objects were fully rendered [27].

A total of 16 able bodied actors within the age range of 19 to 46 participated in the test. None had undertaken any prior training on the production system, or had any background in TV. All were given the same instructions and presented with a comprehensive set of sample tasks for training purposes. The actors were informed that the nature of the task was to assess how well they were able to keep their

hands on the sides of the object while carrying out the required interaction. They were instructed to perform the interaction task to the best of their ability, as a real actor would. For each participant, the experiment was formed of a 10 minute training/instruction routine, a 5 minute break/question period and around 25 minutes for the experiment proper. Each subject filled out a standardised consent form, participants were given no compensation and all data collected was anonymised.

Each subject was positioned so their hands were at a distance of 2 meters from the feedback monitor, camera and the Kinect. Each interaction task required the actor to move a virtual object using two hands from a starting location for a distance of 78.6 cm while viewing the composited scene on the feedback monitor. The interaction was made possible by tracking the position of the actor's hands and locking the centroid of the virtual object to the midpoint of the actor's hands.

From a study of potential motion capture techniques, the Microsoft Kinect [28] interfaced with OpenNI [29] proved to be a convenient, and sufficiently accurate, method of motion capture for measuring the location of an actor's palms. This was aided by the optimal operating conditions of the Kinect being congruent to the experiment design (i.e. where the user stands upright  $\sim 2$  m away from the sensor unit, under controlled lighting conditions). The accuracy and precision of the Kinect was validated by calibrating it to a visual marker tracking system [30].

### 3.2 Analysis

We report on the 72 interaction tasks (trials) that each actor was required to complete, representing every permutation of the conditions in Table 1. These were presented to each actor in a uniform pseudo-random order.

Two key Performance metrics are used to assess the quality of the interaction. Mean Distance to Object Surface (MDOS) (1) is a measurement of how close to the desired location on the surface the actor places their hands, where at all times the object is constrained to be at the midpoint between the hands. The Variability in Distance Between Hands (VDBH) (2) indicates the amount of instability the actor has in hand placement during the interaction

$$MDOS = \frac{1}{N} \sum_{i=1}^N (R_i - L_i) - S, \quad (1)$$

where  $N$  represents the number of video frames captured during the interactive task,  $L$  and  $R$  are the location of the Left and Right hands of the actor, and  $S$  denotes the length of the virtual object along the axis of the hands. Thus MDOS gives the average of the sum of the distance between each hand and the corresponding object surface during the task. In our case, because the object is constrained to the midpoint between the actor's hands, MDOS will be equal to double the average distance of each hand to the surface

$$VDBH = \sqrt{\frac{1}{N} \sum_{i=1}^N ((R_i - L_i) - S)^2}, \quad (2)$$

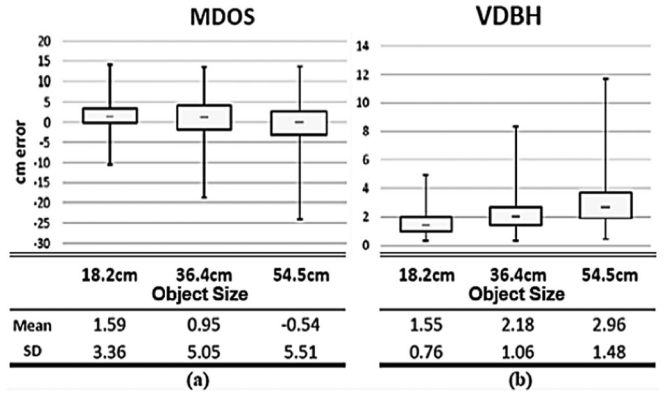


Fig. 2. MDOS (a) and VDBH (b) for each Size condition (These boxplots show the Median, Inter-Quartile ranges and the Max/Min result).

where VDBH represents 1 standard deviation in the distance of the hands of the actor from the object surface during the task (for the case of symmetrical hand placements). The requirement is for this condition to be as low as possible with a VDBH of zero indicating that the subject retained constant hand separation during the interaction.

Upon collection of the data, we observed that occasionally extreme outliers were present. In these cases, it is probable that the actor created a large unintentional motion during the interaction, which could interfere with the interpretation of the results. Therefore, the removal of outliers beyond 3 standard deviations is conducted for each task, which ensured that only the extreme outliers were removed (equivalent to  $\sim 1$  percent of the observations).

Statistical significance is assessed by using a repeated measures Analysis Of Variance (ANOVA) with an alpha of 5 percent in a  $3 \times 3 \times 2 \times 2 \times 2$  design (reflecting the conditions presented in Table 1). A Greenhouse-Geisser correction for violations of sphericity is performed where necessary (represented by ' $\wedge$ '). Post-hoc analysis was conducted in the form of a Tukey test, with an alpha of 5 percent.

### 3.3 Results and Discussion

1) *Effect of object size.* The three object sizes used in this experiment were selected so that any proportional effects can be assessed. Smaller objects ( $< 18.2$  cm) were deemed to be more suited for single handed grasping interactions, while larger objects ( $> 54.5$  cm) would normally require two person interactions. The results are presented in Fig 2 with both the MDOS and VDBH given for the three object sizes.

**Mean distance to object surface**—The results presented in Fig. 2a show that a lower MDOS is present with larger object sizes ( $F_{(2,8)} = 4.327$ ,  $p = 0.027$ ), which decreases from 1.59 to  $-0.54$  cm between the 18.2 and 54.5 cm conditions. Conversely, the standard deviation across the subjects increases from 3.36 to 5.51 cm as larger objects are used.

We observe that as a reflection of the reducing mean with larger object sizes, a shift in the results occurs where a greater proportion of the actors tend to underestimate the size of the object; such that around 25 percent will have underestimated for the smallest object rising to 50 percent for the largest object. This is further reflected in the minimum and maximum recorded results for each task, where

TABLE 2  
Post-Hoc Analysis for the Size Conditions

	MDOS		VDBH	
	P	Sig?	P	Sig?
18.2 cm - 36.4 cm	0.03	Yes	<0.001	Yes
36.4 cm - 54.5 cm	0.478	No	<0.001	Yes
18.2 cm - 54.5 cm	0.017	Yes	<0.001	Yes

the maximum results for each condition are comparable, but the minimum recorded results become appreciably lower with larger object sizes.

Post-hoc analysis showed that there was no significant difference in MDOS between 18.2-36.4 cm, but there was a significant difference between the 36.4-54.5 cm and 18.2-54.5 cm conditions, where an increase is present in both.

**Variability in distance between hands**—The results presented in Fig. 2b show that larger object sizes result in more variability in distance between the hands, ranging from 1.55 for 18.2cm condition to 2.96 for the 54.5 cm condition ( $F_{(1,307, 14,38)}^{\wedge} = 47.162, p = <0.001$ ). Post-hoc analysis showed that a significant effect was present between all conditions (all  $<0.001$ ), with the larger object sizes causing a higher mean in each case. Results for Post-Hoc analysis are presented in Table 2.

2) *Effect of object speed.* In this experiment three target interaction speeds are selected to represent a typical range expected within a real-time interactive MR system. The target speed is communicated to the actors via a dot that moves across the feedback monitor at the desired interaction speed. The actual (mean) speeds that the actors moved the object at were  $0.19 \pm 0.03$  m/s (for the 0.17 m/s target speed),  $0.25 \pm 0.02$  m/s (for the 0.26 m/s target speed) and  $0.37 \pm 0.03$  m/s (for the 0.34 m/s target speed). Results are presented in Fig 3.

**Mean distance to object surface**—The results presented in Fig 3a shows that target speed generally has little effect on MDOS. The mean results of 0.52 cm for 0.17 m/s, 0.67 cm for 0.26 m/s and 0.81 cm for 0.34 m/s are not found to be statistically significant ( $F_{(2,8)} = 0.526, p = 0.599$ ). As is the standard deviation, shown here across the different target speed conditions, to be comparably similar therefore showing little change for all three interaction speeds.

**Variability in distance between hands**—The results presented in Fig 3b illustrate how speed of interaction did not impact greatly on the VDBH, resulting in no statistical significant difference across the mean results for all subjects ( $F_{(2,8)} = 0.007, p = 0.993$ ). The results again show a comparable standard deviation was observed across the conditions.

3) *Effect of other single conditions.* For the Axis of Motion condition, analysis of the MDOS showed no statistical significance ( $F_{(1,10)} = 0.005, p = 0.944$ ), however the VDBH does show a statistical significance ( $F_{(1,11)} = 6.407, p = 0.028$ ), which manifests as an increase of 0.35 cm in the variability between actors hands for the Vertical axis (Vertical = 2.41 cm, Horizontal = 2.06 cm).

No statistically significant results were found for the actor Hand Placement (MDOS:  $F_{(1,10)} = 3.178, p = 0.105$ ;

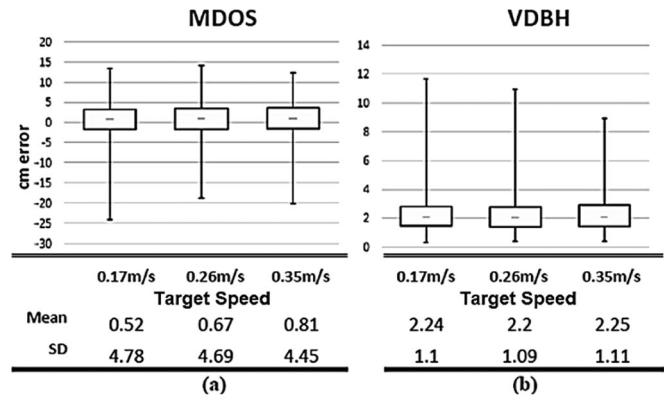


Fig. 3. MDOS (a) and VDBH (b) for each Target Speed condition.

VDBH,  $F_{(1,11)} = 0.172, p = 0.686$ ) or Direction conditions (MDOS:  $F_{(1,10)} = 0.107, p = 0.751$ ; VDBH,  $F_{(1,11)} = 1.727, p = 0.216$ ).

4) *Effects of main interactions between conditions.* When considering the interaction between conditions, the following statistically significant effects were found. For MDOS an interaction was found between the Size and Axis of Motion conditions ( $F_{(2,20)} = 6.899, p = 0.005$ ). Consistent with the Size condition, considered singly, as larger virtual objects are used the MDOS decreases along both axes. However, this is more extreme along the Horizontal axis, with a decrease from 2.1 cm (18.2 cm) to -1.44 cm (54.5 cm). Along the vertical axis MDOS only decreases from 1.07 cm (18.2 cm) to 0.4 cm (54.5 cm). Post-hoc analysis found that the difference between size conditions on both axes was significant between 18.2 and 54.5 cm. The only statistically significant difference between the Horizontal and Vertical axes was for the 54.5 cm object ( $p = 0.006$ ).

Other MDOS interactions between conditions were detected. For Size and Speed ( $F_{(4,40)} = 2.795, p = 0.039$ ) with the 54.5 cm object there was a slight trend towards a larger MDOS, ranging from -0.98 to 0.17 cm with faster interaction speeds. For Speed and Direction ( $F_{(2,20)} = 4.792, p = 0.020$ ) at faster interaction speeds a slight trend towards a lower MDOS was observed when the object was moved to the right (from 1.56 to 0.71 cm) and a higher MDOS when the object was moved upwards (ranging from 0.64 to 1.77 cm).

*Effect of hand position.* ‘Hand Position’, referring to the interaction between the axis of hand placement and axis of the motion conditions, was found to be particularly interesting. The two axis of motion conditions are along the horizontal path of interaction (H) and the vertical path of interaction (V), with the hand placement conditions described as Left/Right (LR) and Top/Bottom (TB). With this notation, to indicate a horizontal interaction path with a hand placement on the top and bottom “H-TB” is used. As illustrated in Fig. 4. We can describe hand position as being parallel to the axis of motion (H-LR and V-TB) or perpendicular to the axis of motion (H-TB and V-LR).

**Mean distance to object surface**—With regards to the hand position of the actors, a statistically significant effect was detected ( $F_{(1,4)} = 5.611, p = 0.039$ ). As shown in Fig. 5a the V-LR condition holds the largest MDOS at 1.83 cm, although this condition also yields the lowest standard deviation (4.09 cm) of all the conditions. From this, we

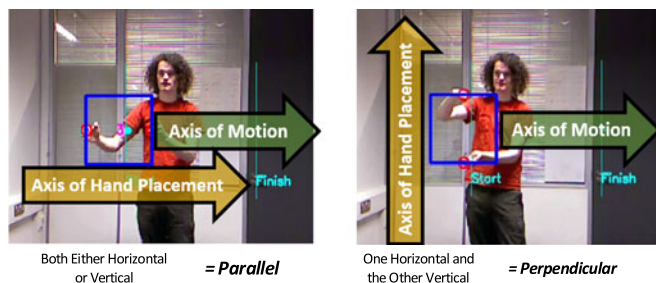


Fig. 4. Hand Position, showing examples of hand placement ‘Parallel’ (H-LR and V-TB) and ‘Perpendicular’ (H-TB and V-LR) to the axis of motion.

observe that lower MDOSs’ are accompanied by higher standard deviations.

Post-hoc analysis showed that there is only a significant difference between the V-LR condition and the other conditions, where an increase in mean distance between hands was found.

**Variability in distance between hands**—The results presented in Fig. 5b show that improved performance is seen when the hands are positioned perpendicular to the axis of object motion. These statistically significant results ( $F_{(1,4)} = 35.139$ ,  $p = < 0.001$ ) show that a lower mean VDBH was observed in both the V-LR (1.76 cm) and H-TB (1.36 cm) conditions than the V-TB (3.06 cm) and H-LR (2.75) conditions. This is reflected in the standard deviations, where the values for the perpendicular conditions are also lower than for the parallel.

Post-hoc analysis demonstrated that the only non-significant difference between the conditions was between the H-LR and V-TB (both ‘parallel’) conditions, which were both larger than the perpendicular conditions. A significant difference was detected between the perpendicular conditions, H-TB and V-LR, but the mean difference was comparatively small. The full results for Post-Hoc analysis are presented in Table 3.

**Summary**—As statistically significant results were found for both the Object Size and Hand Placement conditions for the MDOS and VDBH, the null hypothesis that the conditions tested do not affect actor performance is rejected. An actor’s hand position on the object surface during a bimanual interaction is affected by the Object Size and the hand position relative to the axis of motion.

## 4 VIEWER PERCEPTION OF ERRORS

### 4.1 Background

The results from Section 3 highlight some of the problems actors experience when interacting bimanually with virtual

TABLE 3  
Post-Hoc Analysis for Hand Placement Conditions

	MDOS		VDBH	
	P	Sig?	P	Sig?
HLR-HTB	0.998	No	<0.001	Yes
HLR-VLR	0.010	Yes	<0.001	Yes
HLR-VTB	>0.999	No	0.351	No
HTB-VLR	0.018	Yes	0.009	Yes
HTB-VTB	>0.999	No	<0.001	Yes
VLR-VTB	0.019	Yes	<0.001	Yes

objects, notably errors in both hand placement and placement variance. To quantify the impact these errors would have on a third person (home) viewer, we propose a method for assessing the plausibility of the interaction. We therefore test the null hypothesis that “The measured actor errors from bimanual hand placement will have no perceptual effect on the plausibility of the scene as measured by third party viewers.”

The methodology employed is based on the ITU-R BT.500 recommendations [20] detailed within Section 2. Here a group of observers are presented with a series of videos, with each individual video (known as a video segment) depicting an interaction between an actor and a virtual object with simulated errors based on those measured within Section 3. As we are measuring the perceptual plausibility of the interaction within this experiment, the videos are required to be broadcast quality rendered outputs for all tests, thus it is not possible to directly use the interaction feedback videos detailed in Section 3 as they do not represent the standard of a typical broadcast system. Furthermore, the use of videos depicting a controlled set of conditions allows the effect of each condition to be individually assessed allowing us to test the impact of each measured error independently.

### 4.2 Experimental Methodology

The series of video sequences presented to the observers depict an actor moving a virtual object. The axis of motion and hand position are both horizontal (H-LR). H-LR was identified as a common interaction hand placement from our case study analysis of interaction in an IVS, hence its selection for this study. In each video segment produced, the virtual object (a virtual cube with a wooden crate texture as shown in Fig 6.) followed the centroid of the actor’s hands throughout the video as a real object would. For all presented video segments, a static distance between the actor’s hands and the object is maintained throughout the interaction. The video footage was recorded in a H.264 format in a professional level film studio with appropriate lighting. In each video segment, the actor performs the required motion with a fixed distance between their hands being controlled using a telescopic apparatus set to remain at a defined length. The apparatus was removed from the video in post-production without leaving any significant visual artefacts. The actor in each video knew the objectives of the videos and was given instructions on the interactions required. A single recording is made for each object size moving at the set speed, with the spacing of the hands set at the sum of the box size and the MDOS for that condition.

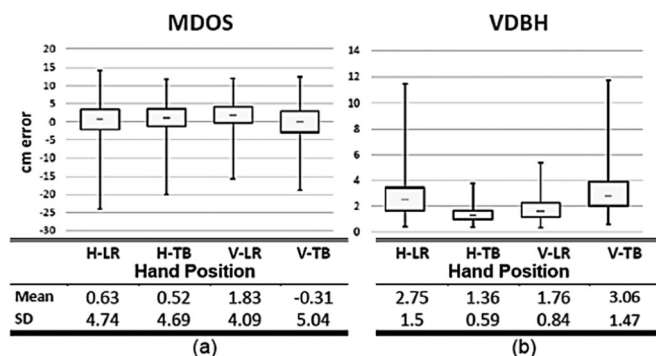


Fig. 5. MDOS (a) and VDBH (b) for each Hand Placement condition.

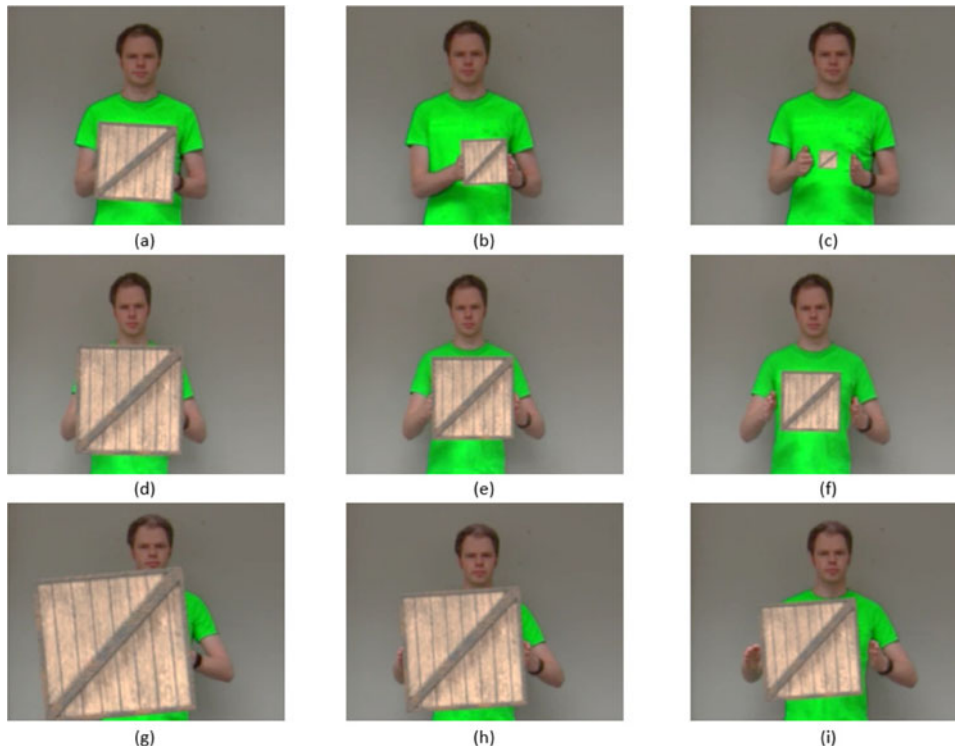


Fig. 6. Screen shots from the perception experiment. (a) to (c) are from the 50px hand distance videos (MDOS errors are (a) =  $-30\text{px}$  (underestimation): (b) =  $0\text{px}$  (perfect fit): (c) =  $+30\text{px}$  (overestimation)), (d) to (f) are from the 100px hand distance videos (MDOS errors are (d) =  $-30\text{px}$ : (e) =  $0\text{px}$ : (f) =  $+30\text{px}$ ), (g) to (i) are from the 150px hand distance videos (MDOS errors are (g) =  $-30\text{px}$ : (h) =  $0\text{px}$ : (i) =  $+30\text{px}$ ).

Adobe After Effects CS5 was used to add the virtual object between the hands. Here the size of the object is adjusted to reproduce each error condition as specified in Table 4. This is done because recording multiple videos with different hand distances and fixed object sizes was found to introduce many inconsistencies between video sequences (e.g., inconsistent speed or angle). Using a single recorded video with one hand distance and with multiple object sizes provides the same MDOS error, but does not have these issues. Thus 43 video segments are produced based on three original recordings. Before editing, the footage is cropped and down sampled to fit the specifications of the monitor. See Fig 6 for images from these videos.

In each video segment, a controlled level of MDOS error was replicated and observers are asked to rate how perceptible this error is on a scale of 1 to 5. The Mean Opinion Score (MOS) from viewing each error condition is analysed alongside the results of the motion experiment to reveal if the errors expected to be commonly introduced would be problematic in real applications. Within this work, we only

analyse the effect of the MDOS of the three size conditions from Section 3, 18.2 (here 50px), 36.4 (here 100px) and 54.5 cm (here 150px), since this was found to make the amount of distinctive data trends explicit and the analysis manageable. Table 4 contains the conditions presented in this study.

*Observers*—A total of 22 observers participated. All had at least an undergraduate education, typically from a computer science background. None were experts in video analysis and none had undertaken any similar visual perception tests previously.

Each observer's visual acuity was measured using a Snellen chart (where 1 is equivalent to 20/20 vision), with an average left eye acuity of 0.92 and right eye acuity of 0.89 being found; the observers used any visual aid they would typically wear for viewing television. No observer with a visual acuity of  $<0.8$  was accepted. Each observer was also required to pass an Ishihara test to exclude for colour blindness.

*Layout*—The experiment is conducted in a low luminance environment using a 14 inch calibrated JVC (TM-H140PN) CRT video monitor adjusted to the recommended specifications as stated by BT.500. The approximate distance between each observer and the video monitor was set at the Preferred Viewing Distance of 150 cm, as defined in the BT.500.

Observers were required to rate each video using the standard five discrete quality categories described adjacently. These categories can then be represented numerically. Before the experiment, the following instructions were given to the observers: "You are about to see a sequence of videos that each depict an actor moving a

TABLE 4  
Range of Replicated Errors Presented in the Experiment

Amount Error (px)	Underestimation					Overestimation					
	-30	-20	-15	-10	-5	0	+5	+10	+15	+20	+30
Object Size (px)		65	60	55	50	45	40	35	30		
	130	120	115	110	105	100	95	90	85	80	70
		180	170	165	160	155	150	145	140	135	130

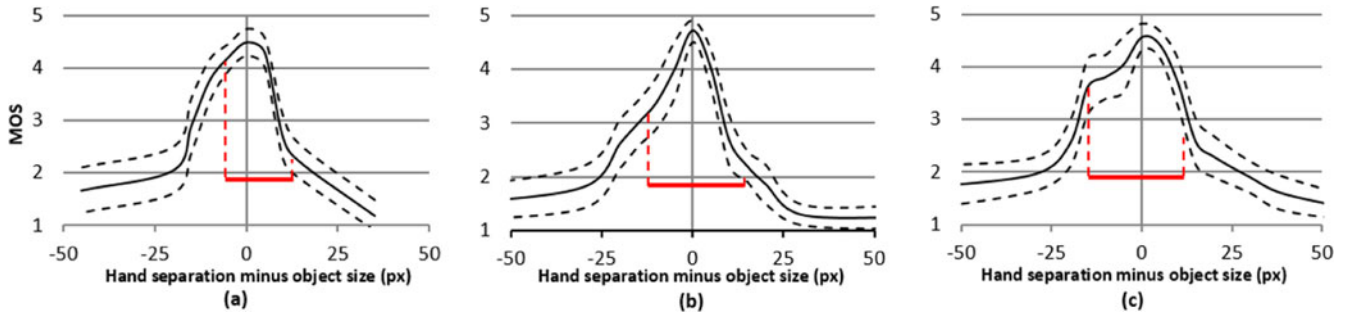


Fig. 7. Results from the perception experiment. The solid lines show the MOS for the 50px condition (a), the 100px condition (b) and the 150px condition (c). The dashed lines show the confidence interval to 95 percent. The red bars represent 95 percent of the actor motion error for the respective size condition.

virtual box. You will be asked to rate how plausible each of the interactions seem compared to those you’ve seen in real life. You will rate them using 5 = Imperceptible (error); 4 = Perceptible, but not annoying; 3 = Slightly Annoying; 2 = Annoying; 1 = Very Annoying.”

*Video presentation*—The observers were presented with the video segments in a Single Stimulus format. This method shows each observer a series of video segments, each presented once. They were presented to the observers in the following order:

1. A training sequence of seven video segments. This is in compliance with the BT.500 and allows observers to familiarise themselves with the range of content they will be shown. It also provides an opportunity to ask any questions.
2. Presentation of video segments for subjective assessment, with a 3 second blank (voting space) between each. The video segments were presented in the following order:
  - i. six video segments encompassing the range of conditions for the test. This allows for stabilisation of observers’ results and can be discarded from further analysis.
  - ii. Presentation of 60 video segments used for the experiment (30 representing the conditions of this study and 30 for the studies presented in Section 5).

The range of selected conditions are derived from the results of the size conditions in Section 3, replicating MDOS errors of up to two standard deviations away from the mean for each object size. Additionally, two anchor videos (an overestimation and underestimation to 4SD) and one reference video depicting ideal hand placement, are also produced for each size condition.

In Section 3 we presented the results in terms of real-life distance measurements (cm) as these reflect properties of human variability in physical interaction irrespective of the recording method, however in the perception studies the viewers will perceive these distance errors as pixels on the screen, which can also be quantified as angles of arc subtended at the eye. For the recording set-up of Section 3, 1cm is equivalent to 2.75 pixels and for the viewing setup in this Section 1 pixel is equivalent to a viewing angle of 1.08 arcmin.

For consistency with the results presented in Section 3, the conditions presented in Table 4 reflect the *total* distance

between the hands and the object surface. This is to say, that if the total distance between the hands is +30, this would be the sum of two +15 errors between each hand and the nearest surface (Fig 7c)

### 4.3 Results

Results from each condition are presented using Mean Opinion Score, which is the average quality assessment of a video segment taken across a number of observers (calculated using Equation 3). The MOS ranges from 1 (Very Annoying) to 5 (Imperceptible). Graphically each MOS is presented with a 95 percent confidence interval

$$\bar{u}_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk}, \quad (3)$$

where N is the number of observers after the removal of outliers and  $u_{ijk}$  is the individual opinion score, with i representing the observer, j representing the object size and k is the error condition.

A Friedman’s test using an alpha of 5 percent is conducted for each of the object sizes. Post-hoc analysis between the conditions is performed using a Wilcoxon signed ranks test, with significance levels of 0.179 percent for the 50 px object size and 0.091 percent for the 100 px/150 px object sizes derived from a Bonferroni adjustment.

The results shown in Fig. 7 represent the 50px, 100px and 150px studies. For each study as the simulated MDOS condition becomes more extreme, the MOS degrades (50px:  $\chi^2_{(9,N=22)} = 134.76$ ;  $p < 0.001$ , 100px:  $\chi^2_{(11,N=22)} = 181.625$ ,  $p < 0.001$ ; 150px:  $\chi^2_{(11,N=22)} = 177.425$ ,  $p < 0.001$ ).

A notable feature of each study is the asymmetry of the results. The results show that when the actor underestimates the size of a straight sided virtual object (i.e. the virtual box) a gradual MOS degradation is detected, however, as they overestimate the virtual object size, a sharp drop in MOS is experienced between approximately the +5px to the +10px conditions.

Post-hoc analysis indicated that the sudden degradation is statistically significant between the 5px and 10px overestimation examples for the 50px and 100px objects (50px:  $Z = -3.776$ ,  $p < 0.001$ ; 100px:  $Z = -3.553$ ,  $p < 0.001$ ) and between the 5px and 15px conditions for 150px object ( $Z = -3.981$ ,  $p < 0.001$ ).



TABLE 5  
MOS Values at Intervals of the Motion Result

MDOS interval	Mean-1SD	MDOS	Mean+1SD
18.2 cm (50px)	4.2	4.4	2.3
36.4 cm (100px)	3.3	4.5	2.2
54.5 cm (150px)	3.8	4.5	2.7

This demonstrates that observers have a low tolerance towards the gap, between the actor’s hands and the virtual object surface, which appears when the actor overestimates the virtual object size. The null hypothesis that “the measured actor errors from bimanual hand placement will have no perceptual effect on the plausibility of the scene as measured by the third party viewers’ is therefore rejected.

The results of this perceptual study were compared to MDOS results of the motion analysis experiment. Table 5 presents the MOSs for each object size condition at the point of the MDOS, 1 Standard Deviation below (underestimation) and 1 Standard Deviation above (overestimation) the MDOS.

It is observed that at 1 standard deviation below the MDOS for each object size yields an MOS between 3.3 (100px) to 4.2 (50px), which is perceptible but not overly distracting. Consequently the tendency of actors to underestimate the size of a large virtual object as discussed in Section 3 is not as important as it initially appears, as the viewers are relatively tolerant of this error. However, due to the asymmetry of the perceptual results at one standard deviation above the MDOS the corresponding MOS score is low. It was found to range between 2.2 (100px) to 2.7 (150px), which means the viewers would find it annoying. This outcome demonstrates that the chance of an overestimation occurring that is distracting to a viewer is likely. However the asymmetry found within the results could be related to both the form of interaction and the type of virtual object used within the tests. O’Sullivan and Dingliana [31] identified that viewers are better able to spot discrepancies in collisions between spherical objects than between complex shaped objects. Therefore analysis of objects with curved surfaces and with other interactions may yield different results.

The tests were limited to the case of H-LR hand placement with a square object. However, relevant features of human perception of motion, such as different interaction paths, have been identified previously that could guide conditions in future experiments. Reitsma and Pollard [32] identified that ballistic motion is perceived differently on the horizontal and vertical axes, with errors in vertical motion being easier to mitigate. Fuller and Carrasco [33] have shown that humans are more perceptive of incorrect object motion trajectories in the vertical axis than the horizontal. Thus, the results from the current work should not be directly generalised to other motion trajectories; these need to be explored in further work.

From our overall findings, it is possible to conclude that an actor is more likely to underestimate the size of a large object, and thus horizontal bimanual interactions involving larger objects are less likely to have a perceptual impact on a viewing audience. Alternatively it may be necessary to

train actors manipulating smaller objects such that they consciously bias to underestimate the object size.

This conclusion demonstrates that basing design decisions on an analysis of motion alone is inadequate, as this would have predicted the employment of smaller objects to reduce MDOS error.

## 5 SCENE ADAPTATION TO MITIGATE ERROR

To mitigate the impact of the object size estimation errors noted in Section 3, and to aid in improving the MOS score for a third party viewer of the interaction from Section 4, two adaptation solutions are proposed and evaluated. The first adapts the size of the virtual object based on the measured actor errors. The second adapts the real scene properties, namely the colour of the exposed gap between the actor and the virtual object in an attempt to conceal the errors. These solutions were identified as being potentially useful or interesting in a pilot study with six participants. Both of these solutions are evaluated using the same BT.500 recommendations detailed in Section 4.

The 30 videos for the conditions presented in this part of the study were shown incognito within the experiment discussed in Section 4 (which we now term as the ‘static size study’). The participants were not informed of the nature of these videos.

### 5.1 Adaptation of Virtual Object Size

Through analysis of interaction within the virtual TV studio we found that frequently at the start of the interaction actors failed to correctly locate the surface of the virtual object. In addition, during interaction their hand placement would vary. To overcome this, we propose a method of object adaptation which initially allows the object size to grow to match the hand placement error (representing a tween over a number of frames) and then adapts the size immediately to any further change within the hand position throughout the interaction (allowing a constant matched adaptation).

These two methods are represented here using separate sets of videos so that the impact on plausibility for each one can be determined individually. Two size conditions are used for each method: a 100px and a 150px hand distance. The telescopic apparatus controlled the hand placement growth over the 1.4s video length.

*Tweened adaptation*—the virtual object adapts from an initial size to fit the distance between the actor’s hands. As identified in Section 3 overestimation was found to be a more common type of error than underestimation. When actors interact with virtual objects they often have difficulty initially locating the precise position of the object surface. In a practical mixed reality system this could take an unreasonable amount of time if spatial tolerances are not applied. Consequently this study explores adapting a virtual object to the distance between the actor’s hand location starting from an initial overestimation. The conditions replicated for the tweened adaptation study in 10 video segments are based on the same overestimation conditions as in Section 4 so that a direct comparison can be made. Images from the tweened adaptation videos are presented in Fig. 8.

*Matched adaptation*—the virtual object continuously adapts to the varying distance between the actor’s hands.

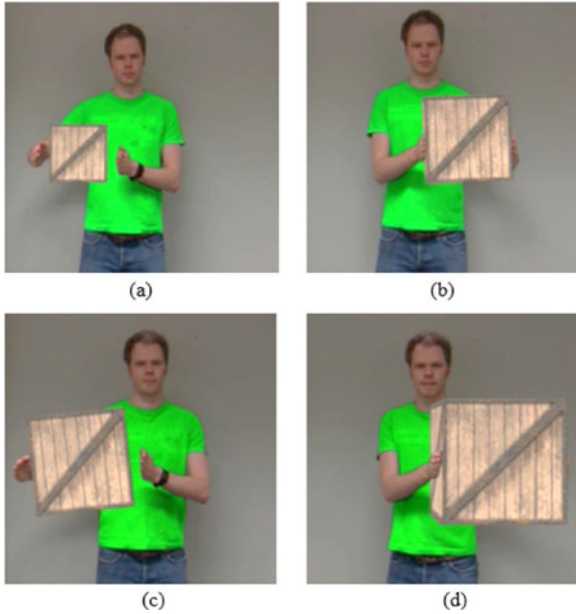


Fig. 8. Screen captures from tweened adaptation videos. Images (a) and (b) show the tween adaptation in the 100px hand distance and (c) and (d) showing the 150px hand distance, each case presenting an object growth of 30px.

As discussed in Section 3, when the actor completes an interaction the distance between their hands will also vary, either becoming further apart or closer together. To continue examining the effects of overestimation, these 11 videos explore cases where the actor's hands drift further apart. Images from the matched adaptation videos are presented in Fig. 9. Here the actor is correctly estimating the object size at the start of the interaction – this could be assumed to be the case if the tweened adaptation had already occurred or if the actor already had their hands placed accurately on the sides of the virtual object.

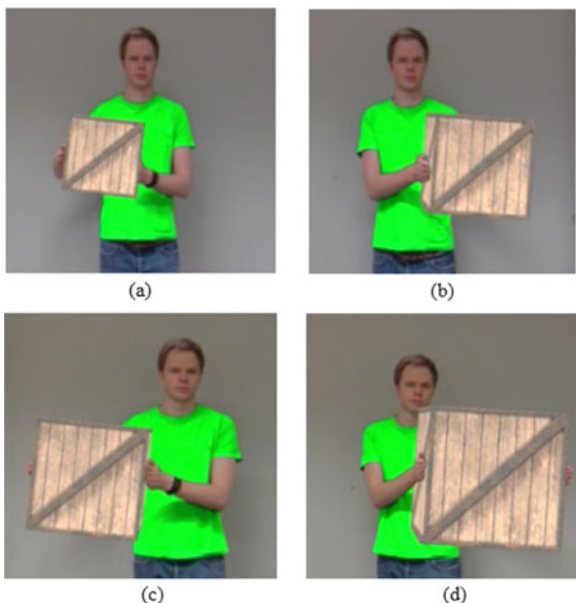


Fig. 9. Screen captures from matched adaptation videos. Images (a) and (b) shows 70px to 100px and (c) and (d) showing 120px object size expanding to 150px matching the actor's hands, each case presenting a growth of 30px.

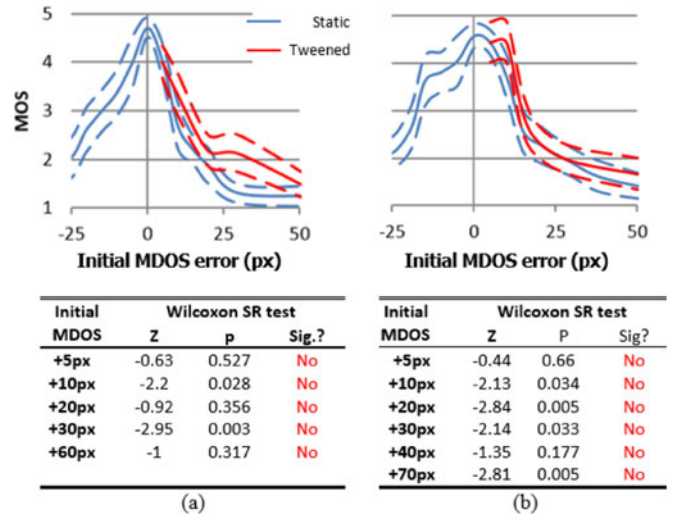


Fig. 10. Results from the tweened adaptation conditions compared to the equivalent static size conditions. (a) shows the MOS for the 100px hand distance and (b) shows 150px. Red lines represent the tweened adaptation object, blue represents the static size object (dashed lines show 95 percent confidence interval).

When testing for statistical significance between the static size object and the adapted object (tweened and matched), a standard two-way repeated measures ANOVA would not provide a valid result due to the format of the data collected in this experiment being non-parametric and ordinal. Instead statistical analysis will be performed as a series of individual 2-tailed Wilcoxon signed ranks tests with an alpha of 5 percent, using a pairwise comparison between the corresponding conditions of each study.

### 5.1.1 Tweened Adaptation—Results

The goal of this experiment was to determine whether the process of adapting the size of the object to match the actor's hands gave a greater plausibility to the interaction than the static error conditions in Section 4

Fig. 10 presents the comparison between the results from Section 4 (the static size) and tweened adaptation studies for objects finishing with sizes of (a) 100px and (b) 150px, alongside the results of the pairwise Wilcoxon Signed Ranks test. Here the 0px MDOS error represents the ground truth condition.

A Bonferroni adjustment provided a new significance level of 0.0011 for the 100px hand distance condition and 0.00076 for the 150px hand distance condition. For both object size conditions the MOS trend of the tweened adaptation study closely matched that of the static size study and no statistically significant effect was measured between any of the corresponding conditions.

The effect that the speed of object adaptation had on the plausibility of an interaction was also investigated. A series of four additional videos depicting an initial size adaptation over five different time periods were also assessed. This study used a 150px hand distance size and a starting MDOS of +20px, representing an overestimation.

Fig. 11. presents the results for the speed of adaptation. This illustrates how the MOS ranged between 3.28 when the adaptation lasted 0.16 s and 2.72 for an adaptation lasting 1.42 s, therefore giving a difference of only 0.56.

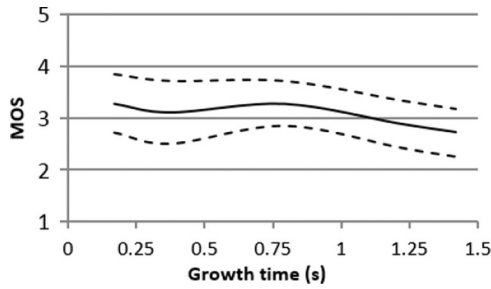


Fig. 11. MOS scores for the tweened adaptation of the virtual object from a starting MDOS of +20px to a final MDOS of 0px over a series of time periods from 0.16 to 1.42 s (dashed lines show 95 percent confidence interval).

The result from the Friedman’s test determined that a statistically significant effect could be present ( $\chi^2 = 9.469$ ,  $p = 0.05$ ). Post-hoc analysis in the form of a two-tailed Wilcoxon signed-ranks test was then conducted between each pair, with a Bonferroni adjustment providing a new significance level of  $p < 0.005$ . No statistically significant difference between any conditions was detected. It was also noted that after the tests no observers commented positively or negatively on the adaptation speed.

### 5.1.2 Matched Adaptation—Results

The results of this experiment are shown in Fig. 12. A Bonferroni adjustment provided a new significance level of 0.0033 for the 100px hand distance and 0.0017 for the 150px hand distance. No statistically significant difference was found to exist between the matched adaptation and static size techniques until the sudden drop, that was observed in the static size study, occurs (+10px for the 100px hand distance and +15px for the 150px hand distance). After this point, the results became statistically significant and the matched adaptation technique consistently outperformed the static size technique. The benefit this technique offered was considerable with a 2.27 and 1.77 improvement in MOS for an error of +30px with the 100px and 150px hand distance studies respectively, effectively eliminating the sudden degradation observed in the static size study.

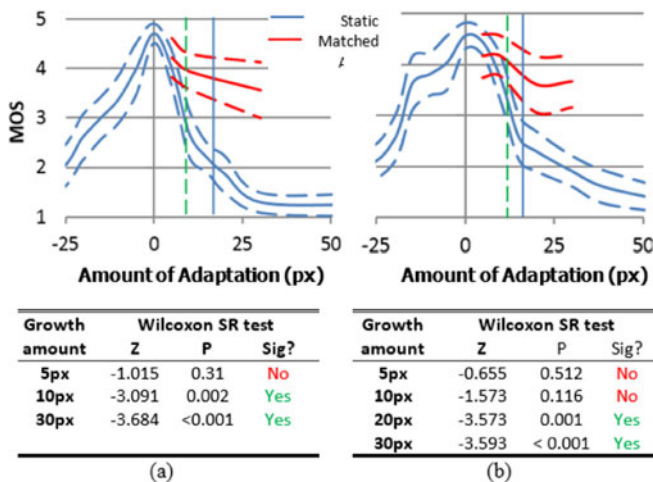


Fig 12. Results from the matched adaptation conditions compared to the equivalent static size conditions. (a) Shows the results for the 100px hand distance and (b) the 150px hand distance. Red lines represent the matched adaptation condition, blue represent static (dashed lines show 95 percent confidence interval).

It is important to note that during the post-test debriefing six out of the 22 observers reported they were aware of the growth of the virtual object—but assumed it was because the actor was moving the virtual object towards the camera slightly, instead of the virtual object growing in size itself. In effect, this natural response of humans to assume that when an object appears to change size it is moving towards or away from them has helped to enhance the plausibility of the interaction.

We can demonstrate the impact offered by the matched adaptation method by relating the results to the observed interaction errors from Section 3. This allows us to measure the improvement this matched adaptation solution offers over the static object results from Section. In Fig. 12 the solid blue vertical lines represent the MDOS overestimation to one standard deviation from the motion analysis experiment (Section 3), for the results of the object size condition. In both cases (Figs. 12a and 12b) adapting the size of the virtual object was shown to mitigate the error effectively at this point. The adaptation technique shows  $\sim 1.4$  MOS improvement for the 100px hand distance size and  $\sim 0.8$  MOS for the 150px hand distance size. The dashed green lines represents 1 standard deviation from the mean for the VDBH. Again it was possible to see that adapting the size of the virtual object had a positive effect at this point. The matched adaptation allowed an MOS improvement of  $\sim 1.2$  for the 100px hand distance and  $\sim 0.7$  for the 150px hand distance.

Due to the viewer’s improved tolerance towards the overestimation that object adaptation allowed, it was possible to determine that the actor’s misestimations could be compensated for by continuously adjusting the size of the virtual object.

## 5.2 Adaptation of Real Scene Properties

The results from Section 4 clearly show that overestimation of object size had a negative effect on the MOS for viewers, with the gap between the actor’s hands and the virtual object causing the lower score. As an alternative to the adaptation of the object size, the goal of this experiment was to quantify whether the background colour within the exposed ‘gap’ region between the actor’s hands and the object could be simply altered to reduce the viewer perception of the error, therefore leading to an improved MOS.

Throughout this part of the study a single 150px hand distance is used with an MDOS of +30px, representing an overestimation of +15px on either size of the object. In the video the background exposed within the ‘gap’ during the interaction is the actor’s shirt. The colour of the shirt (originally green) was adjusted in post-production, using standard chromakey methods, to a range of possible background colours without impacting any other feature of the video. The selected colours were pure colour channels for the case of green, black, blue, red, white, while the actor’s skin colour was sampled from their hand; creating 6 videos in total. For comparison the ground truth video is the original green condition with a 0px MDOS error which scored an MOS of 4.5. As with Section 5.1, the videos for the conditions presented in this study were shown within the experiment discussed in Section 4, therefore forming part of the same 60 video test set.

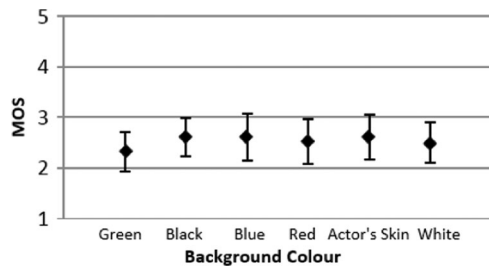


Fig. 13. The effect of background colour on the plausibility of interaction: The MOS for each condition shows that background colour has no significant effect on the plausibility of interaction. The whiskers represent the 95 percent confidence interval.

The results presented in Fig 13 show there was little difference between each condition. A Friedman's test was conducted using an alpha of 5 percent to determine whether there was any statistical significance among the results. The result  $\chi^2 = 3.958$ ,  $p = 0.556$  confirmed that background colour had no statistically significant effect within the range tested. No observers commented affirmatively or negatively either. As such, there was no evidence to show that background colour affects the perception of the MDOS error.

## 6 CONCLUSION AND RECOMMENDATIONS

This paper presented a generalizable framework for measuring the fidelity and plausibility of actor interaction errors in mixed reality systems. The framework comprises three stages. The first involved capturing the interaction error caused by actors performing bimanual interactions, the second stage assessed the perceptual plausibility of this erroneous interaction, and the final stage evaluated three potential error mitigation methods for improving the interaction plausibility.

For the interaction error measurement we assessed the motion of actors when completing a controlled series of bimanual interaction tasks. We defined two performance metrics suitable for this, namely the Mean Distance to Object Surface, which measured the hand placement inaccuracy to the object surface, and the Variability in Distance Between Hands, which measured the extent of variance in an actor's hand position during the interaction.

For the perceptual tests, we assessed the plausibility of the interactions from the perspective of third party screen viewers. Following the ITU-R BT.500 recommendations [20] we defined a series of perceptual tests where viewers were shown a series of broadcast standard videos containing controlled replications of the identified actor interaction errors. Viewers rated the plausibility of each interaction on a five point scale with a Mean Opinion Score determined for each interaction. These tests were limited to the scenario of the actor moving an object horizontally by placing their hands on the left and right sides (H-LR). The perceptual effects of the further scenarios which were part of the motion study will be investigated in future work.

The final part of the framework detailed measures for improving the plausibility of the interaction. We assessed three interaction adaptation methods: firstly tweening the object size to match the MDOS error, secondly immediately adjusting the object size based on any hand placement variation and finally adapting the background colour of any exposed gap region between the actors' hands and the object.

From the findings, the following set of guidelines is given:

- Smaller objects can lead to an increase in the mean distance to the object surface (MDOS) as actors fail to accurately locate the object surface.
- The amount of variability in the distance between the actor's hands (VDBH) can be reduced by having the actor place their hands *perpendicular* to the axis of object motion during the interaction. Using smaller virtual objects can also improve the quality of the interaction by providing a reduced VDBH.
- Overestimation of the object size by the actor of  $>10\text{px}$  is more likely to result in a perceptible error from third person viewers. Therefore encouraging actors to underestimate the object's size is suggested. Automatically adapting the size of the virtual object to match the actor's hands during a bimanual interaction can improve the viewer's plausibility of the interaction. This improvement can mitigate the degradation in interaction plausibility observed for large overestimations of object size.

Application of these guidelines can aid both actors and content producers in developing improved interactions, offering an increased plausibility to the viewer. Actors could be advised to bias towards object underestimation, as they tend to overestimate the size of (smaller) virtual objects, whilst an underestimation error was found to be less perceptible than an equivalent overestimation. In further work we will explore enhanced rendering methods for actor feedback so as provide additional visual and saliency cues with a view to reducing actor interaction errors. For content producers, in the case of an overestimation error, using a virtual object with an adaptable size continuously matching the distance between the actor's hands was found to be an effective technique for mitigating the error.

The success of the overall framework is in its ability to highlight the impacting conditions and then assess techniques to improve interaction plausibility. This framework would be beneficial for researchers studying fidelity in MR or for content developers looking to create a more plausible MR scene. This leads us to propose that with appropriate adjustments the framework is transferable for measuring and assessing realism between real and virtual elements for a range of domains across the MR spectrum.

## REFERENCES

- [1] T. Lee and T. Hollerer, "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking," in *Proc. 11th IEEE Int. Symp. Wearable Comput.*, 2007, pp. 83–90.
- [2] G. Hough, C. Athwal, and I. Williams, "ScaMP: A head guided projection system," *ACM Des. Interactive Syst.*, p. poster/2 page paper, 2012.
- [3] G. Hough, C. Athwal, and I. Williams, "Advanced occlusion handling for virtual studios," in *Proc. 6th Int. Conf. Convergence Hybrid Inf. Technol.*, 2012, vol. 7425, pp. 287–294.
- [4] G. Hough. (2013) [Online]. Available: [www.youtube.com/user/GPHResearch](http://www.youtube.com/user/GPHResearch)
- [5] G. Hough, I. Williams and C. Athwal, "Measurements of live actor motion in mixed reality interaction," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 99–104.
- [6] G. Hough, I. Williams, and C. Athwal, "Measurement of perceptual tolerance for inconsistencies within mixed reality scenes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 343–344.

- [7] M. Minoh, H. Obara, T. Funtatomi, M. Toyoura, and K. Kakusho, "Direct manipulation of 3D virtual objects by actors," in *Proc. 2nd Int. Conf. Inf. Res. Develop. Knowl. Soc. Infrastructure*, 2007, pp. 11–18.
- [8] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Experimental Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.
- [9] D. Bowman, D. Johnson, and L. Hodges, "Testbed evaluation of virtual environment interaction," in *Proc. ACM Symp. Virtual Reality Softw. Technol.*, 1999, pp. 26–33.
- [10] R. Teather and W. Stuerzlinger, "Guidelines for 3d positioning techniques," in *Proc. Conf. Future Play*, 2007, pp. 61–68.
- [11] A. Sutcliffe and B. Gault, "Heuristic evaluation of virtual reality applications," *Interacting Comput.*, vol. 16, no. 4, pp. 831–849, 2004.
- [12] Y. Guiard, "Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model," *J. Motor Behavior*, vol. 19, pp. 486–517, 1987.
- [13] N. Kim, W. Woo, G. Kim, and C.-M. Park, "3D virtual studio for natural inter-acting," *IEEE Trans. Syst., Man Cybern.*, vol. 36, no. 4, pp. 758–773, Jul. 2006.
- [14] B. Woldecke, T. Vierjahn, M. Flasko, J. Herder, and C. Geiger, "Steering actors through a virtual set employing vibrotactile feedback," in *Proc. 3rd Int. Conf. Tangible Embedded Interaction*, 2009, pp. 169–174.
- [15] M. Klapdohr, B. Woldecke, D. Marinos, J. Herder, and C. Geiger, "Vibrotactile pitfalls: Armguidance for moderators in virtual TV studios," in *Proc. 13th Int. Conf. Humans Comput.*, 2010, pp. 72–80.
- [16] M. Haller, "Photorealism or/and nonphotorealism in augmented reality," in *Proc. ACM SIGGRAPH Int. Conf. Virtual Reality Continuum Appl. Ind.*, 2004, pp. 189–196.
- [17] A. Hertzman, "Painterly rendering with curved brush strokes of multiple sizes," in *Proc. 25th Annu. Conf. Comput. Graph. Interactive Tech.*, 1998, pp. 453–460.
- [18] J. Fischer, D. Cunningham, D. Bartz, C. Wallraven, H. Bulhoff, and W. Straßer, "Measuring the discernability of virtual objects in conventional and stylized augmented reality," in *Proc. 12th Eur. Conf. Virtual Environ.*, 2006, pp. 53–61.
- [19] W. Steptoe, S. Julier, and A. Steed, "Presence and discernability in conventional and non-photorealistic immersive virtual reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 213–218.
- [20] ITU-R BT.500–13, "Methodology for the subjective assessment of the quality of television pictures," 2012.
- [21] S. Winkler, A. Sharma, and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in *Proc. Int. Symp. Wireless Multimedia Commun.*, 2001, pp. 547–552.
- [22] Z. Wang, A. Bovik, and B. Evans, "Blind measurement of blocking artifacts in images," in *Proc. Int. Conf. Image Process.*, 2000, vol. 3, pp. 981–984.
- [23] H. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 11, pp. 317–320, Nov. 1997.
- [24] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," *Electron. Imag. Int. Soc. Optics Photon.*, vol. 5292, pp. 182–193, 2004.
- [25] S. Deshpande, "A method for synchronization mismatch perception evaluation for large ultra high resolution tiled displays," in *Proc. Int. Workshop Quality Multimedia Exp.*, 2009, pp. 238–243.
- [26] W. Ijsselsteijn, H. de Riddera, R. Hamberga, D. Bouwhuisa, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," *Displays*, vol. 18, no. 4, pp. 207–214, 1998.
- [27] S. I. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 1993, pp. 488–493.
- [28] Microsoft, Kinect [Online]. Available: <http://www.xbox.com/en-us/live/kinect>, 2014.
- [29] Primesense, OpenNI. (2014) [Online]. Available: [www.openni.org](http://www.openni.org)
- [30] Pixoft, "Vicasso information," Pixoft. (2009) [Online]. Available: <http://www.pixoft.co.uk/Vicinfo.htm>
- [31] C. O'Sullivan and J. Dingliana, "Collisions and perception," *ACM Trans. Graph.*, vol. 20, no. 3, pp. 151–168, 2001.
- [32] P. Reitsma and N. Pollard, "Perceptual metrics for character animation: Sensitivity to errors in ballistic motion," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 537–542, 2003.
- [33] S. Fuller and M. Carrasco, "Perceptual consequences of visual performance fields: The case of the line motion illusion," *J. Vis.*, vol. 9, no. 4, pp. 1–17, 2009.



**Gregory Hough** received the BSc degree in television technology and production in 2009 and the MPhil degree in engineering in 2012. He received the PhD degree at Birmingham City University, where he focused on the creation of an interactive virtual studio environment. His research has explored aspects of actor feedback, real-time occlusion and interaction. Part of this study focuses on the plausibility of interactions from the perspective of a third party viewer.



**Ian Williams** received the BSc degree in media technology and the PhD degree in image processing from Manchester Metropolitan University in 2004 and 2008, respectively. He is currently part of the DMT Lab, Birmingham City University, where he is a senior lecturer in digital technology. His research interests include low-level image processing, 2D edge and 3D surface detection, image feature extraction and manipulation, and also mixed reality interaction.



**Cham Athwal** received the BSc degree from Manchester University, the MSc degree from London University, and the PhD degree from Aston University in Birmingham all in physics. He has held research positions at the Universities of Southampton and Lancaster and at CERN in Geneva, as well as a spell leading an R&D group in industry. Since 1990, he has been on the academic staff at Birmingham City University, where he is currently a professor of digital technology and the head in the Digital Media Technology (DMT) Lab. He has led many industry/academic collaborations developing a range of multimedia technologies, most notably research underpinning current industry standard vehicle crash test analysis systems. His current research is focusing on mixed reality and image and audio processing.