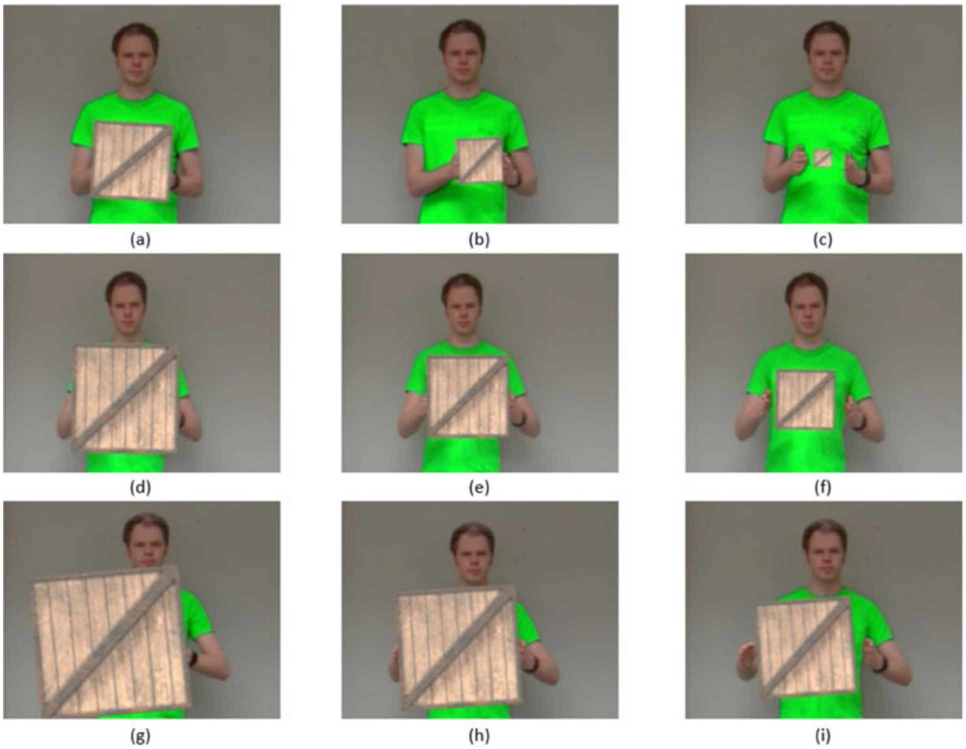


Fig. 6. Screen shots from the perception experiment. (a) to (c) are from the 50px hand distance videos (MDOS errors are (a) = –30px (underestimation): (b) = 0px (perfect fit): (c) = +30px (overestimation)), (d) to (f) are from the 100px hand distance videos (MDOS errors are (d) = –30px: (e) = 0px: (f)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

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.

TABLE 4
Range of Replicated Errors Presented in the Experiment

	Underestimation						Overestimation				
Amount Error (px)	−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	120

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 adjectively. 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

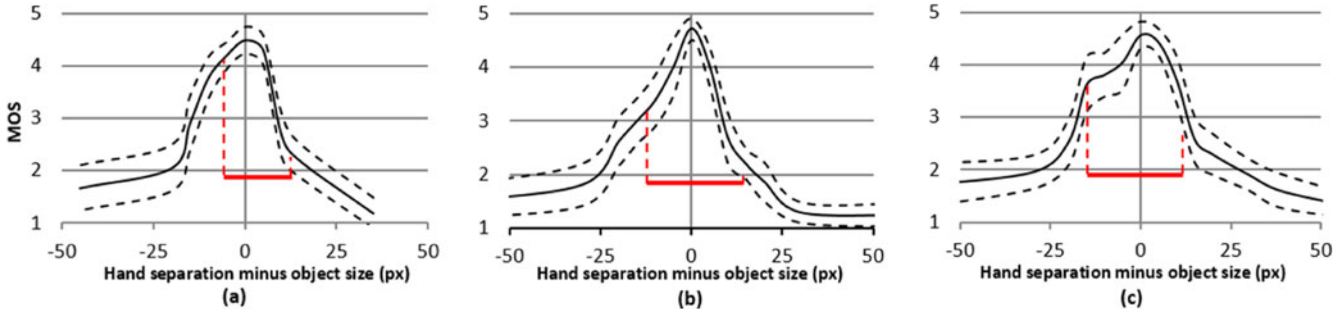


Fig. 7. Results from the perception experiment. The solid lines show the MOS for the 50px condition (a), the 100px condition (b) and the 150px condition (c). The dashed lines show the confidence interval to 95 percent. The red bars represent 95 percent of the actor motion error for the respective size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

$$\bar{u}_{jk} = \frac{1}{N}\sum_{i=l}^{N} u_{ijk}, \tag{3}$$

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.

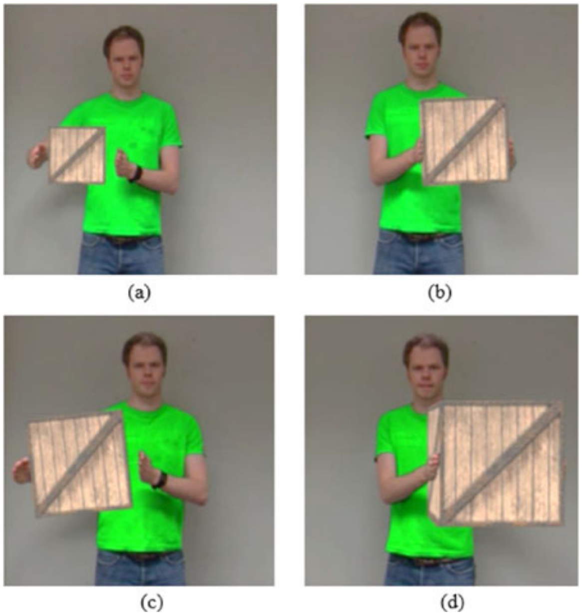


Fig. 8. Screen captures from tweened adaptation videos. Images (a) and (b) show the tween adaptation in the 100px hand distance and (c) and (d) showing the 150px hand distance, each case presenting an object growth of 30px.

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.

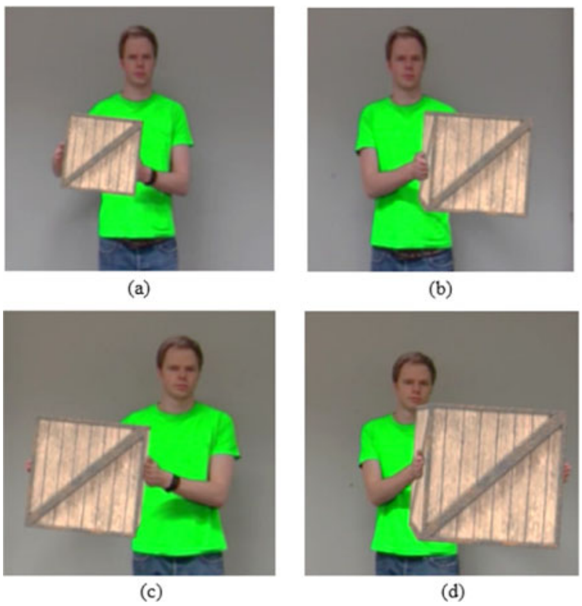
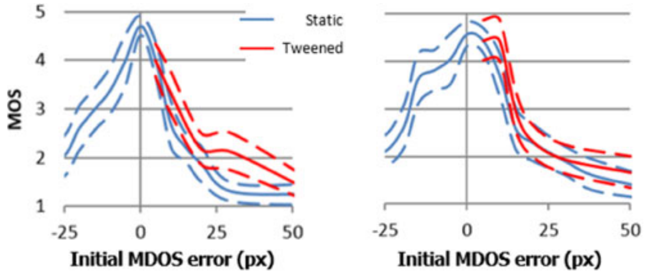


Fig. 9. Screen captures from matched adaptation videos. Images (a) and (b) shows 70px to 100px and (c) and (d) showing 120px object size expanding to 150px matching the actor's hands, each case presenting a growth of 30px.



Initial MDOS	Wilcoxon SR test Z	p	Sig.?
+5px	-0.63	0.527	No
+10px	-2.2	0.028	No
+20px	-0.92	0.356	No
+30px	-2.95	0.003	No
+60px	-1	0.317	No

(a)

Initial MDOS	Wilcoxon SR test Z	P	Sig?
+5px	-0.44	0.66	No
+10px	-2.13	0.034	No
+20px	-2.84	0.005	No
+30px	-2.14	0.033	No
+40px	-1.35	0.177	No
+70px	-2.81	0.005	No

(b)

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 tweend 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.

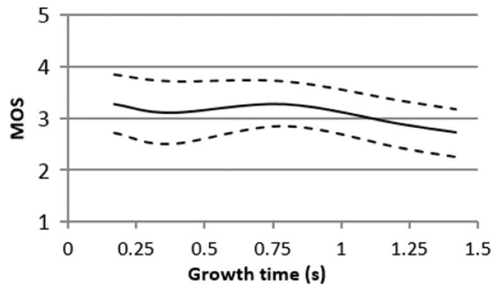
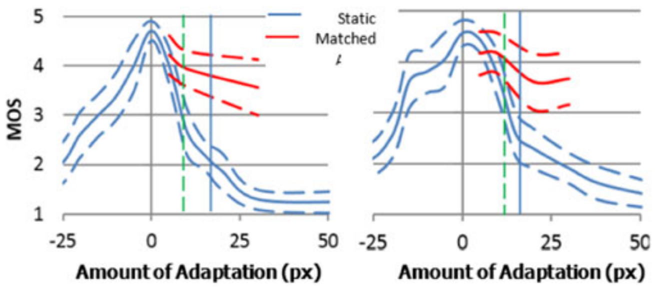


Fig. 11. MOS scores for the tweened adaptation of the virtual object from a starting MDOS of +20px to a final MDOS of 0px over a series of time periods from 0.16 to 1.42 s (dashed lines show 95 percent confidence interval).

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.



Growth amount	Wilcoxon SR test Z	P	Sig?
5px	-1.015	0.31	No
10px	-3.091	0.002	Yes
30px	-3.684	<0.001	Yes

(a)

Growth amount	Wilcoxon SR test Z	P	Sig?
5px	-0.655	0.512	No
10px	-1.573	0.116	No
20px	-3.573	0.001	Yes
30px	-3.593	< 0.001	Yes

(b)

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 ~1.4 MOS improvement for the 100px hand distance size and ~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 ~1.2 for the 100px hand distance and ~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.

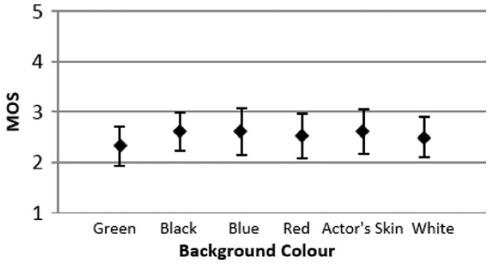


Fig. 13. The effect of background colour on the plausibility of interaction: The MOS for each condition shows that background colour has no significant effect on the plausibility of interaction. The whiskers represent the 95 percent confidence interval.

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 >10px 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
[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: Armguidence 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. Straβer, "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

[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.