This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2017.2658570, IEEE Transactions on Visualization and Computer Graphics

Augmented Reality vs Virtual Reality for 3D Object Manipulation

Max Krichenbauer, Goshiro Yamamoto, *Member, IEEE,* Takafumi Taketom, *Member, IEEE,*
Christian Sandor, *Member, IEEE,* and Hirokazu Kato, *Member, IEEE*

Abstract—Virtual Reality (VR) Head-Mounted Displays (HMDs) are on the verge of becoming commodity hardware available to the average user and feasible to use as a tool for 3D work. Some HMDs include front-facing cameras, enabling Augmented Reality (AR) functionality. Apart from avoiding collisions with the environment, interaction with virtual objects may also be affected by seeing the real environment. However, whether these effects are positive or negative has not yet been studied extensively. For most tasks it is unknown whether AR has any advantage over VR.

In this work we present the results of a user study in which we compared user performance measured in task completion time on a 9 degrees of freedom object selection and transformation task performed either in AR or VR, both with a 3D input device and a mouse. Our results show faster task completion time in AR over VR. When using a 3D input device, a purely VR environment increased task completion time by 22.5% on average compared to AR ($p < 0.024$). Surprisingly, a similar effect occurred when using a mouse: users were about 17.3% slower in VR than in AR ($p < 0.04$). Mouse and 3D input device produced similar task completion times in each condition (AR or VR) respectively. We further found no differences in reported comfort.

Index Terms—H.5.1 Information Interfaces and Presentation - Multimedia Information Systems - Artificial, Augmented, and Virtual Realities; I.3.6 Computer Graphics - Methodology and Techniques - Interaction techniques

◆

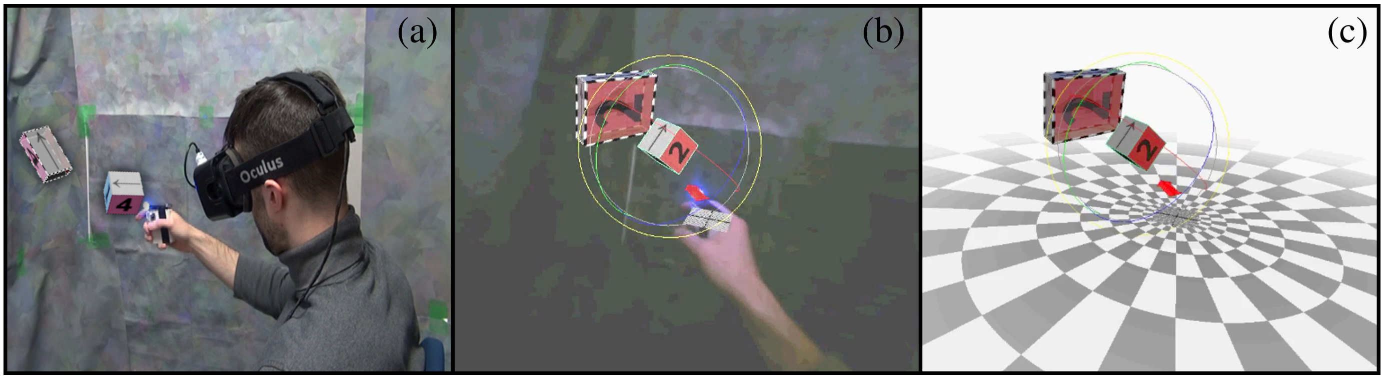


Fig. 1: Our study compared task performance in Augmented Reality (AR) versus Virtual Reality (VR) on an object selection and transformation task. (a) Illustration of the AR condition with 3D input device. (b) View through the head-mounted display (HMD) in the AR condition. (c) View through the HMD in the VR condition.

1 INTRODUCTION

Both AR and VR User Interfaces (UIs) have so far been used for a great number of tasks, where they at times have shown great promise for increasing a user's performance compared to traditional mouse-and-monitor UIs. However, usually the tasks that have been studied differ greatly due to the different focus of both technologies: usually, AR studies involve interaction with real-world objects that could not be performed in purely virtual environments.

Our hypothesis was that seeing the real environment (in AR) has a significant effect on task performance even when working with virtual objects, due to a more direct understanding of spatial relations. Prior research has not produced conclusive evidence whether the ability to see the real environment (in AR) has any effect on task performance or user satisfaction. To our knowledge, we present the first study to directly compare AR and VR in a classical 3D object selection and placement task setting. 3D object placement is a very general task that has possible implications on task performance in almost all 3D interaction tasks.

We asked participants to perform the same task with both a 6DOF 3D input device (Figure 1(a)) and a traditional 2D computer mouse (Figure 2), in both AR and VR (Figure 1(b) and (c)). The task consisted of selecting and transforming a "source" object to resemble a "goal" object in position, orientation, and scale in three dimensions (9DOF; Figure 3). We included the mouse condition to further shed light on the possible reasons for performance differences in AR and VR. When using the 3D input device, the user's hand and the virtual cursor are perfectly aligned, providing additional visual feedback to 3D interactions.

Fig. 2: Illustration of a user in the mouse condition.

In the mouse condition, the only additional feedback was seeing an empty work area, which should not provide much benefits on task performance other than a general sense of orientation, spatial limits to movement, and possibly a sense of connectedness with the real world. On the other hand, these positive effects might be counterbalanced by increased sensory load in AR vs. VR, reducing or even reversing the overall effect. Therefore, our experiment consisted of four conditions: AR with 3D input device, VR with 3D input device, AR with mouse, and VR with mouse. In each condition, we asked participants about their subjective level of comfort in order to find whether people preferred AR or VR, or if using a 3D input device resulted in increased strain from arm motions in mid-air.

Our results show a statistically significant increase in performance in AR over VR when a 6DOF 3D input device is used: in the VR environment, it took participants on average almost 22% more time to perform the task. To our own surprise, we found a similar, yet reduced effect when participants used the mouse: about 12% reduction of task performance in VR compared to AR. While most participants expressed a preference for either mouse or 3D input device, we could not find a statistically significant overall trend of either device being perceived as more comfortable to use.

Contribution:

- Our work is the first to give quantitative evidence on whether AR work environments can outperform VR work environments in 3D interaction tasks.
- Because we held our task fairly general, our results can be applied to other similar applications where 3D object selection and transformation is required. This may include games, Computer Aided Design (CAD), or even training simulations. This provides both manufacturers and researchers with an incentive to further probe the capabilities of AR.
- Our results furthermore create new research questions on what causes the respective effects found in the mouse condition, where the visual stimuli of AR did not provide any immediately apparent benefits. This in turn brings into question how much of the observed performance improvements with a 3D input device are due to the commonly cited effects of visual feedback from seeing ones' own body, and how much are intrinsic to other factors of AR.

2 RELATED WORK

Only little prior work has been published that directly compares AR and VR in the same setting.

Boud et al. [1] compared several systems as a training tool for assembly tasks, including VR and AR settings. They find AR to outperform several VR variants, which in turn outperform conventional instructions by a great margin. However, their AR system was what they describe as "context-free". This means the AR graphics were not registered with the real world and could be described as a Heads-Up Display (HUD) of a static diagram image. The measured performance was in the assembly of real objects, for which the various AR and VR conditions were only used as prior training for the task. Our work focuses on 3D interaction tasks with virtual objects in AR or VR.

Jones et al. [2] compared AR and VR systems in their effect on human depth perception using an HMD. Based on prior work that had shown that depth is consistently underestimated by users in a VR setting, they tested whether a similar effect existed in AR. Their results showed that no such underestimation of depth occurs in AR. This indicates that the additional spatial cues of the real environment may help spatial understanding. However, no interaction with virtual objects was required in their task. Furthermore, they only analyzed depth estimation at a range of 2-8m, which is far beyond the usual work area for most tasks.

Cidota et al. [3] performed a similar study with a focus on serious games. They measured subjective usability and task performance in AR and VR under various visual effects such as blurring and fading on a simple hand-based select-and-move task designed to measure depth perception performance. The pair-wise within-subject comparisons found no statistically significant effect in mean task performance for neither the different visual effects nor when comparing AR and VR. Only when they removed all data except those participants who got their best score in only one sub-condition they found significant differences. However, whether AR or VR performed better depended on the visual effect and partially contradict their results on measured performance.

Juan and Pérez [4] studied differences between AR and VR in acrophobic scenarios. Instead of performing a task, the objective was to expose participants to anxiety or phobia provoking situations. They find both AR and VR to be effective for creating anxiety at appropriate moments in the simulation but find no statistically significant advantage of either AR nor VR.

Arino et al. [5] compared AR and VR directly using an autostereoscopic display instead of an HMD. Their participants were children (8 to 10 years) and the task they performed resembled more of a game in which children were asked to passively count specific objects in the scene. The children could only interact with the scene by rotating a single fiducial marker around one axis, which would rotate the virtual object on it in order to see it from a different angle. They did not find significant differences in the mean task completion time, nor in post-use questionnaires regarding the experience. However, AR was generally preferred by the children over VR in direct comparison.

Botden et al. [6] compared two systems for laparoscopic surgery simulation, the LapSim VR and ProMIS AR. The AR system was found to be more realistic, to have better haptic feedback, and to be more useful for training purposes. However, this study mostly describes differences between two competing systems. This does not necessarily imply general differences between AR and VR, since a better VR simulator could easily be build by improving haptics and so on.

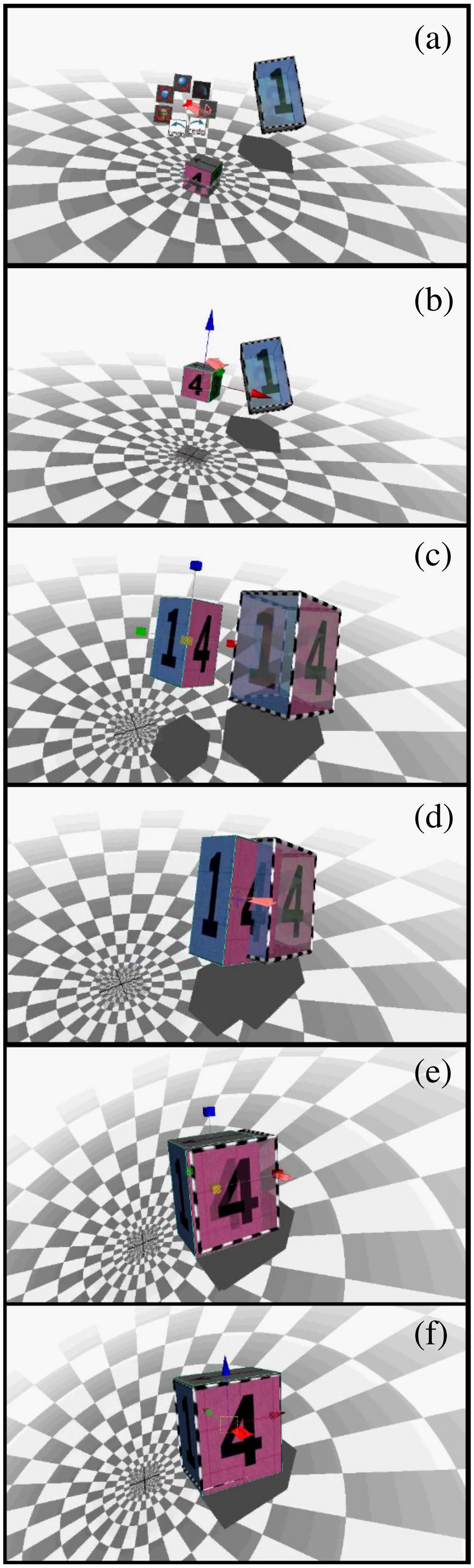


Fig. 3: The task in the VR + 3D Input Device condition, as seen through the HMD. (a) The 3D scene at the beginning with the source object in the center, and the tool menu visible on the left. (b) Translating the object. (c) Scaling the object in 3DOF after rotation. (d) Moving the object into the goal. (e) A look from the side reveals that the object length is not correct. (f) The object lies within tolerance thresholds. The task is completed.

Sandor et al. [7] performed a user study on object selection performance in both AR and a simulated half-mirror VR condition. They simulated a mirror-based VR system by displaying a virtual semi-transparent screen floating over the work area on a video see-through (VST) HMD. Comparing user performance on an object selection task, they find that the AR condition was superior to the simulated VR condition. However, their conditions differed in several factors such as head-tracking, object visibility and occlusion. Therefore, the results do not necessarily indicate a general advantage of AR over VR in all settings.

Irawati et al. [8] created a 3D edutainment environment for falling domino blocks, which can be used both as an AR and a VR environment. However, they did not perform any evaluation on advantages or disadvantages of either method.

Similarly, Rhienmora et al. [9] created a simulator for dental surgery that can be used either in AR or VR. A preliminary evaluation by an expert dental instructor indicated that the AR version better resembles a real clinical setting. However, only the AR version was used with an HMD, while the VR version was displayed on a 2D screen. Furthermore, no quantitative evaluation was performed.

Lee et al. [10] investigated the effect of visual realism of the environment on task performance in a search task, by comparing an AR system with VR simulations with varying degrees of rendering realism. They recreated a real outdoor environment and found some indication that visual simplification in level of detail, texture, and lighting may have some positive effect on task performance. However, their virtual environment differed from the augmented environment in several areas, such as added objects and changes in vegetation. Most of the performed tasks did not show significant differences between AR and VR performance.

Möller et al. [11] performed a user study on indoor navigation with an AR guidance system on a hand-held device, which had an alternative “VR” mode. They found that users navigated a path about 15% faster with VR than with AR. Furthermore, VR appeared to be more robust to errors than AR. However, the VR mode differed greatly from the AR mode. In the VR mode, the device could be held at a low angle and allowed manually changing the view direction of pre-recorded panorama images in a drag-and-pan fashion, resembling more of a panorama picture viewer than actual VR. This may have had a great influence on the result, as participants reported that holding the device upright in their field of view (in AR mode) felt both straining and awkward (since they were worried about the opinion of passers-by).

Khademi et al. [12] compare projector-based tabletop AR with non-immersive monoscopic screen based VR regarding performance on a “pick and place” task designed for rehabilitation of stroke patients. They used healthy subjects for their evaluation and found that they performed better in the AR condition than in VR. In both conditions, they interacted with a physical object. The AR or VR gear was only used to display target object placement areas, which means that in the VR condition participants had to perform a mental transformation from the on-screen computer graphics (where they only saw the placement object and target area, not their own hand) to the table surface where they had to perform the task.

Bowman et al. [13] explored the possibilities of using a high-end VR system to simulate different VR and AR systems of equal or lower display fidelity in VR in order to analyze the effects of specific design factors of those systems such as field of view (FOV) or latency. They argue that comparisons between

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2017.2658570, IEEE Transactions on Visualization and Computer Graphics

existing VR and AR systems are inherently ambiguous because any two systems will differ in a number of factors such as FOV, weight, form factor, and so on, making it impossible to isolate the effects on any one single factor. They further provide evidence for the viability of their approach by recreating prior studies with real systems in their simulation and achieving similar results. However, they admit problems due to the technical limitations of the simulator such as delay and lack of photorealistic rendering capabilities in VR. We go the opposite way by using an AR system to simulate VR by artificially blocking out the live video stream. This removes the limitations of not being able to simulate AR sufficiently while staying true to the concept of simulating one technology with another in order to isolate certain factors for analysis.

Howlett et al. [14] analyzed differences in task performance and eye movements on the same object sorting task performed in a real environment and a VR reproduction of the same environment using back-projection and a haptic input device. They found that their VR reproduction of reality — while quite close to the original — had some effects on the participants. In VR, people took longer to perform the task, had a longer average eye fixation duration, and tended not to look ahead to plan their next steps. The average saccade amplitude was similar in VR and reality within each task. This points to the possibility that even slight deviations in reproducing a real environment in VR can have a significant effect. However, since the study compared two groups of only four participants (between-subject), and no statistical analysis of the measurements was performed, the results may not be generalizable to other applications.

Werkhoven and Groen [15] performed a study on task performance as measured in both speed and accuracy on an object placement task in VR, comparing hand-tracking to a table-bound input device. Unlike this study, the mouse was a 3D SpaceMouse. They found that the correctly aligned virtual hand input metaphor performed significantly better, but admit that this may have been influenced by technical factors and task design. In their study, the task was strictly divided into separate sub-tasks, first rotating the object before positioning it. It thus did not resemble a natural work-flow where rotation and translation are used together or in alternation to zero-in on the desired result. Furthermore, the hardware used in their study appears quite outdated by today's standards, and therefore results may differ when using modern hardware.

One of the key advantages of AR over VR is the ability to see one's own body. Regarding the foundational research on visuo-spatial perception, Coello [16] reviewed a large number of publications showing the importance of different depth cues such as one's own limbs or textured backgrounds for correct spatial understanding. Graziano et al. [17] performed neurological experiments on monkeys identifying key areas of visual perception related to seeing one's own arms.

To our knowledge, no prior publication compares interaction performance in AR and VR directly, using the exact same device, set-up, and task. By doing so, we isolate the key factor (the ability to see the real environment) and can provide quantitative evidence for its effect on task performance.

3 USER STUDY

We theorized that being able to view the real environment could have several effects on task performance, which could be either advantageous or detrimental.

The most important effect is that the visual feedback of seeing one's own hand is often considered helpful to perform tasks that require some form of hand-eye coordination. In an AR environment however, the key features of this feedback system such as occlusions, shadows, or direct and indirect lighting effects may not be presented physically correct. This could result in confusing the user instead of improving performance. In our study, we ignored all of these factors and rendered the virtual content in both conditions exactly the same in both the AR and VR condition.

Another effect is that seeing the boundaries of the work area may give users more confidence to move around freely and swiftly, without worrying about bumping into physical objects. Furthermore, it may provide a better sense of direction and reduce disorientation, which may be helpful to some users by reducing confusion and cyber-sickness. Again, these could also turn out negatively, when VR causes users to become more daring in their motions and less distracted due to the removal of the real environment.

For the VR condition, it is obvious that the type of the virtual environment affects task performance. It is therefore important to carefully consider the strategy from which to conduct the analysis. The options can be categorized into *(A)* attempts to approximate the real environment in VR; *(B)* creating a fictitious environment; and *(C)* providing as little environmental cues as possible. The same categories naturally correlate to different approaches to representing the user's body in VR, where one can *(A)* attempt to capture detailed information on the user's body and represent it faithfully; *(B)* generate a virtual body based on sparse information (such as the position of the 3D input device); or *(C)* omitting any display of the user's body, only providing a virtual cursor to indicate the device position.

Option *(A)* is usually connected to additional efforts, since common VR or AR HMDs do not provide out-of-the-box solutions to scan and track the environment and the user's body. It is not well researched what efforts end-users are typically willing to take in order to improve their VR experience. Furthermore, it also raises concerns towards the degree of fidelity since the stated end-goal of approach *(A)* is the elimination of differences between AR and VR. Practically, a Video See-Through (VST) AR system that uses depth reconstruction can be technically seen as a VR environment created from ad-hoc reconstruction, because the image and depth buffers are just one form of 3D representation of the environment created from video images which is then used in the rendering process. In such a perfect VR reconstruction of reality, AR and VR become synonymous and any difference found between AR and VR conditions can be seen as a failure to successfully recreate the real environment. While it is interesting to study which failures result in performance differences and which are tolerable, such an analysis is application specific and may be better suited to be performed from a framework of diminished reality.

Option *(B)* appears to be the most practical since it requires no additional effort from the user. However, it bears the risk that certain design choices unduly affect performance measurements. We could no longer be certain that the same task performed in a different virtual environment would not produce different results. For example, the performance of claustrophobic users may decline in larger environments while nyctalopic users may perform better in brighter environments, unrelated to whether the environment is virtual or real. Similarly, hand-tracking may actually decrease performance if precision and reliability fall below a certain threshold. Since common VR hardware only tracks the position of the controllers, the position of the elbow and shoulders must be guessed completely.

Option *(C)* is the most abstract approach. It therefore has a greater potential to yield reproducible results, but may not allow conclusions to be transferable to real-life use cases. However, when we look at common current VR design applications such a Google TiltBrush, Oculus Medium, Adobe Project Dali, or the Unreal Engine VR Editor, we see that neither of them makes an effort to provide an artificial virtual environment other than a horizon, nor do they attempt to display the user's hand's and arms.

We therefore decided to follow approach *(C)* to display a completely empty virtual environment, since it not only allows us to avoid unwanted influences on the result from particular design choices but also is an acceptable approximation of common current real-life applications. We decided to display only a ground plane in order to provide a sense of orientation and limited depth cues, similar to the real environment in the AR condition. Only a simple arrow is rendered to indicate the position of the 3D cursor in the VR condition. We decided not to depart from this purely abstract representation to avoid any advantage that the user might gain from seeing virtual hands. With this, we match current VR design applications such as Google Tilt Brush [1], Oculus Medium [2], Kodon [3], and Tvori [4], which do not show the users hands but only an abstraction of the input device. This allowed us measure the full effect to which visual impressions (including hands, arms, and environment) can positively impact user performance. In interpreting the results of our study it is important to note that the more information we have about the real environment, we more we can start to imitate the AR condition by displaying objects of the real world in the VR condition. For now, we ignore hardware differences such as factors in Optical See-Through (OST) HMDs or latency which are arbitrary to the used devices and change rapidly with every generation. Then we can assume that the closer we resemble the AR condition in VR by providing visual feedback about the environment and the users own body, we would achieve more and more similar results. Thus our study can be seen as a measurement of the maximum difference between VR without any knowledge of the real world, and AR, which is a full visual representation of the real world. Of course, we do not know whether any measured difference between the conditions will disappear gradually as we add more visual information. It is possible that even showing one single polygon indicating the opposing wall could lead to exactly equal task performance in AR and VR. However, since our VR condition is based strictly on the information available to VR developers, our findings are of interest. That is to say that VR developers cannot indicate the opposing wall because they have no information where the opposing wall is. As discussed before, one may choose to "stimulate" possible advantages of the AR condition by guessing the position of walls, objects, or the user's limbs. But for our study, we abstained from simulating arbitrary guesswork, as there is no accepted framework for estimating details about real-world VR installations.

In order to not only find supporting or contradicting evidence to all of these factors combined (AR vs. VR), but also shed some light on what single effects may have an influence on task performance, we decided to perform exactly the same task with a 3D 6DOF input device as well as a traditional 2D mouse. In the case of the mouse, seeing the real environment should not have the same effect, because the cursor is two-dimensional and not directly aligned with one's real hand (which is below the work area and therefore usually outside the field of view). Only the general effect of the virtual environment, including artificial lighting, emptiness and thus, possibly, a sense of isolation and disorientation could affect the user.

Based on the design of our study we formulated 3 hypotheses:

H_1: **Task performance with a 6DOF input device will be significantly improved when the user is able to see the real environment (AR vs. VR).** If this hypothesis is supported, it provides an argument for attaching cameras to VR HMDs in order to increase users performance when using 3D input devices (such as Oculus Touch[5] or HTC Vive controllers[6]) or hand tracking (such as Leap Motion[7]).

H_2: **There will be no difference in performance between AR and VR when using a 2D mouse**. If this hypothesis would be falsified, then there is a possibility that the effects related to H_1 are at least in part dependent on a general sense of spatial awareness instead of task specific visual feedback. For example, purely virtual environments might improve performance by reducing visual cognitive load.

H_3: **Subjective measures of comfort will differ between AR and VR environments, as well as between 2D mouse and 3D input device.** We hypothesized that the isolation of VR environments may have a negative effect on the users' comfort, because the lack of visual feedback of the boundaries of the work area and one's own body may induce a feeling of unease. Supporting evidence for this hypothesis could, on one hand, be seen as an explanation in the case that H_2 is rejected, and on the other hand provide an alternative incentive for using either AR or VR, which is not related to task performance but to user satisfaction. A similar argument could be made regarding the mouse vs input device conditions, and whether users actually want to use a 3D input device solely for its novelty factor.

Although our study design could also allow us to compare the relative performance of the mouse and the tangible 3D input device, we cannot derive useful general information from these measurements. Several arguments can be made regarding the various advantages and disadvantages of each device. Some stress the advantages of 2D mice, like high familiarity of most users, and its stability and precision when placed on a flat surface. Others favor 3D input devices for their higher number of DOF or the correct spatial correlation of the device and the cursor in space. However, because many technical or experimental factors contribute to overall performance of either device, a direct comparison is difficult. Our 3D input device was a prototype, which we specifically designed for this user study, and our results may not apply to other devices.

3.1 Experimental Platform

To test our hypotheses we developed a prototype 3D modeling UI. In order to stay closely related to our target field of application of 3D design, we based our UI on Autodesk Maya[8], by developing it

---

[1] https://www.tiltbrush.com

[2] https://www.oculus.com/medium/

[3] http://store.steampowered.com/app/479010

[4] http://www.tvori.co/

[5] www.oculus.com/en-us/touch/

[6] www.htcvive.com

[7] www.leapmotion.com

[8] www.autodesk.com/maya/

as a plug-in for this software.

Our test system consisted of a Dell Precision Notebook with an Intel Core i5 CPU with 2.90 GHz, 8GB RAM, and a Nvidia Quadro K4100M graphics adapter. The computer was running Windows 7 and Maya 2014.

We used an Oculus Rift DK2 with ovrVision stereo cameras attached to its front to turn it into a VST-HMD. The cameras have a resolution of 800×600 pixels and a frame rate of 25 fps. The cameras were running in both the AR and VR condition and the exact same image processing (image rectification and undistortion) was applied, in order to achieve the same computational load and avoid one condition performing faster than the other. The only difference was that in the VR condition, the prototype cleared the frame buffer instead of using the image from the camera as a background for rendering.

The mouse used was a standard cable-bound laser mouse (DELL MOC5UO). When the mouse was used, a cursor was displayed in the dominant eye view only. The UI was identical to Autodesk Maya with two exceptions: a markup-menu that appeared upon pressing the right mouse button to select the tool (translation, rotation, and scaling, as well as an “undo” and “redo” button), and viewpoint navigation by dragging with the middle mouse button and using the mouse wheel. When the middle mouse button was pressed, the mouse controlled the viewpoint in a tumbling motion around the selected object. The mouse wheel allowed moving forward and backward. We used this style of navigation, which is similar to the camera motion used in Maya and other modeling software products, in order to keep our results closely related to real-world applications.

We used a custom-made 3D printed ergonomic case in the shape of a pistol grip as our 3D input device. Four buttons were attached to the device. The first button was the main interaction button (similar to the left mouse button). Two of the other buttons would bring up the tool menu (similar to the right mouse button in the mouse condition). We used two buttons only for convenience since some test users found one button location easier to reach than the other. The last button on the device was a navigation button (similar to the middle mouse button in the mouse condition) which allowed changing the virtual cameras viewpoint in a “grabbing-the-air” navigation fashion without editing the objects. An Inertial Measurement Unit (IMU) was used to track the orientation of the input device, and an LED was attached to track its position in the HMD-mounted cameras by computer vision and triangulation. This meant that the device would only work when the user was looking at it. It was not possible to use it outside the field of view. The combination of IMU and LED tracking provided 6 DOF (translation and rotation). An Arduino microcontroller and a Bluetooth modem were used to transfer the IMU data and button interactions to the computer over a wireless connection. When using the input device, a 3D arrow was rendered on top of it to clarify where exactly the interaction would take place.

In the VR condition, we displayed a circular pattern to indicate a “ground plane” in order to give the user a basic sense of orientation even when the target objects were not visible. We indicated that this pattern was indeed the “ground level” by displaying simple object shadows on it. In the AR condition, no ground plane was displayed, and therefore no shadows were visible.

The menu and UI was the same for both mouse and 3D input device conditions, except an additional “6DOF Tool” input metaphor which allowed controlling both translation and rotation of the object at the same time, similar to holding a real object.

The work area was about 1×1.5 meters and draped with patterned cloth in order to achieve constant lighting conditions and facilitate tracking. We have chosen our real environment deliberately empty (thus highly similar to the VR environment) in order to measure the effect of AR in and of itself, without possible side effects from helpful or hindering objects in the real environment. Thus we also avoided any possibility for participants to accidentally bump into objects in the VR condition, or distract the user in the AR condition.

In the mouse condition, we placed a plastic board (45×31 cm, 3 mm thick) on the participant’s lap as a surface for the mouse.

3.2 Procedure

Our study included four conditions: using the mouse in an AR setting, using the 3D input device in an AR setting, using the mouse in a VR setting, and using the 3D input device in a VR setting, all of which were performed seated (see Figure 1(a) and Figure 2).

The conditions were performed in a Latin square balanced order. The first time either the mouse or 3D input device were used, a tutorial was displayed in the AR or VR environment that explained the complete set of functions available. This gave the participant some time to practice and ensured that the task was correctly understood. This practice trial was not used in the later task performance analysis.

In the subsequent trials, we measured participants task completion time on a 3D object selection and transformation task of primitive 3D objects. See Figure 3 for an example execution of the task. A textured 3D box and a semi-transparent “goal” object were displayed. The task was to position the “source” object in the same way and at the same place as the “goal” object, by manipulating translation (x, y, z), rotation (yaw, pitch, roll), and scaling in each dimension (width, height, depth). Thus the task required the participant to manipulate the object in 9DOF. The source object was set at the scene origin, aligned with the world coordinate system, and at unit scale at the start of each trial. The goal was positioned at random for each trial with the following constraints: position was always above the ground plane and between 9 and 10 units away from the origin, the scale in each dimension ranged from 0.5 to 3 times the size of the source object. The rotation was randomly chosen without any restriction. The scene was automatically positioned in a 70cm wide (side to side) work area with the source object at the origin in the center, 60cm in front of the user and 35cm below the users’ head. Thus the source object appeared around 4cm in size with the goal around 35cm away. The task was completed when certain precision thresholds were met. These thresholds were 0.15 units in Euclidean distance (in any direction), 8 degrees of rotation (around any vector), and 0.5 units difference in scale (sum of the 3 dimensions of scaling). When all conditions were met a sound would ring to inform participant and experimenter that the task was completed and the next task could be started.

Our sample consisted of 24 volunteer participants (one female, 23 male; ages 22 to 43, average 27.9 years; all right-handed), which were selected among university students and staff members. Before the participants started using the prototype, we determined their ocular dominance with a Miles test (12 right), explained the basic concept of the user study, demonstrated the 3D input device, and provided a tutorial for each UI, which was completed in the

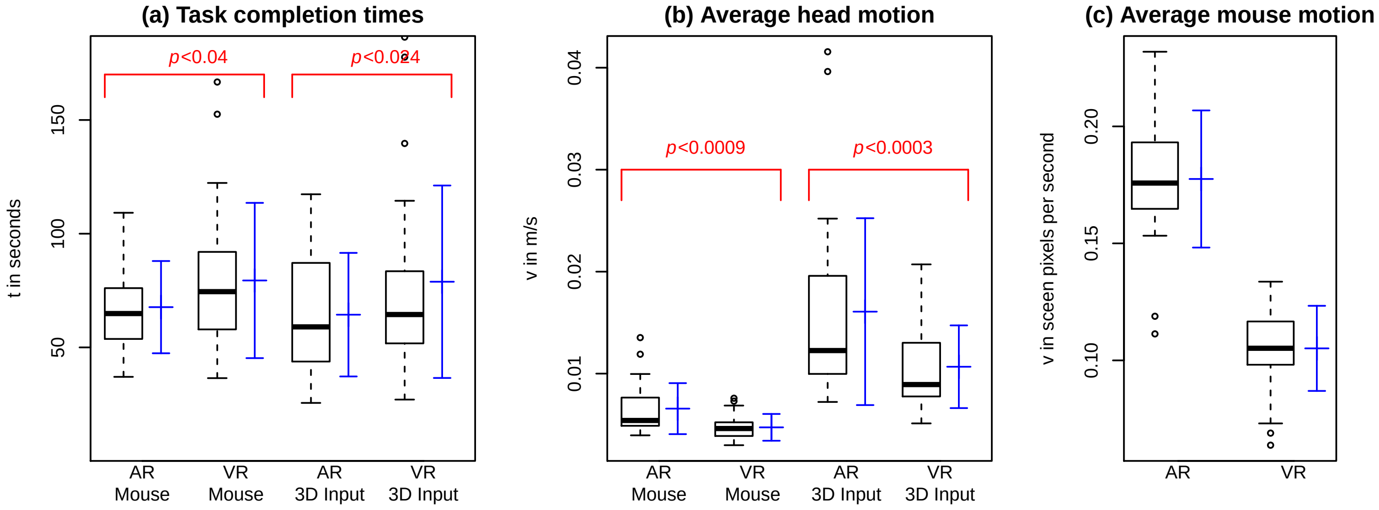


Fig. 4: Experimental results. Blue bars indicate mean and standard deviation. Circles indicate outliers. The *p*-values were obtained by a paired (within-subject) *t*-Test, where each data point is one participants average in that condition.

AR or VR task environment. Before starting each condition, we reminded the participants that time was the critical factor in task completion.

Immediately after the participant had completed the final placement task for each condition, we asked him or her to rate the current feeling of comfort on a scale from one to ten, where one would mean least comfortable and ten would mean most comfortable. These ratings were only intended for a within-subject comparison of relative comfort from one condition to another, so we gave no further guidelines on how to use the rating system.

We recorded six trials per condition for every participant. Throughout the study, we slightly optimized the process of conducting the measurements and fixed technical issues that caused problems, however, the task and criteria always remained the same. The first trial in each condition was discarded as training. Every participant completed all conditions in one session, with one single exception where technical difficulties caused a delay of 30 minutes, in which the participant was allowed to temporarily leave the experimentation area. On several occasions, we encountered problems, either of technical nature or because the participant got confused and reached a state from which he or she could not easily return or reach the goal. In these cases, we reset the condition to a new random state and asked the participant to repeat the task. On some occasions, we accidentally took less than or more than four (non-training) measurements. In order to satisfy all requirements for our statistical analysis strictly, we therefore had to exclude certain data. In those cases where we accidentally took additional measurements, we discarded all measurements after the fourth. In those cases where we did not take enough measurements, we excluded the participant's data from the statistical analysis altogether, which was the case for three participants. Thus the final data set used for the statistical analysis contained 336 measurements from 21 participants.

Since we are working with time as our main metric, the immediate measurements were not producing perfectly normally distributed residuals. In order to meet the requirements for ANOVA, we performed a logarithmic data transformation with base ten [18]. The resulting residuals were approximately normally distributed, which we ascertained with both a Shapiro-Wilk test ($W \approx 0.99, p \approx 0.85$) and an Anderson-Darling test ($A \approx 0.26, p \approx 0.69$), as well visually by generating a histogram and a QQ-Plot (Figure5). We further performed a Barlett's-Test for equal variances ($\chi^2 \approx 3.92, df = 3, p \approx 0.27$).

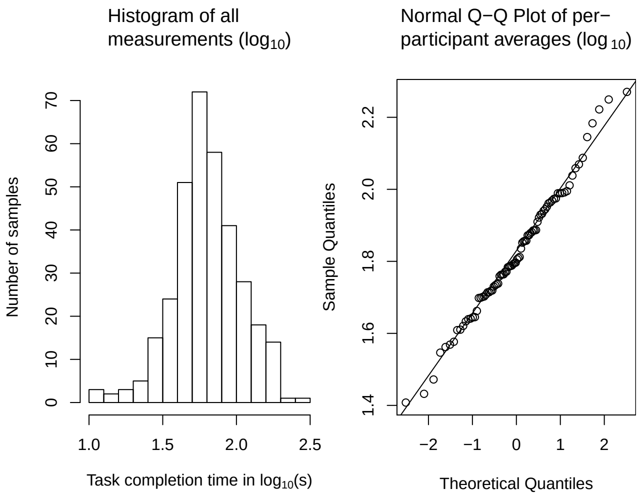


Fig. 5: Normality tests for the resulting data. All data points used are the base ten logarithm of recorded task completion time.

3.3 Results

A summary of the recorded measurements can be seen in Figure 4(a). A two-way repeated measures ANOVA of participant performance showed a significant effect of the environment (AR or VR) with F (1,328) ≈ 4.1, $p < 0.044$ (effect size $\eta^2 \approx 0.023$) but not for the input device used F(1,328) ≈ 0.17, $p > 0.68$). We did not find evidence for a significant interaction effect between the conditions ($F(1,328) \approx 0.4, p > 0.52$; Figure 6).

Thus we have found supporting evidence for H_1, and can accept the hypothesis that AR has beneficial effects on task performance compared to VR when using a 6DOF input device. In the AR condition, average task completion time was reduced by about 14.5 seconds ($\approx 18\%$; $p < 0.024$ on a paired (within-subject) *t*-Test). However, we also found similar (though reduced) effects when participants were using a mouse and therefore have to reject

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2017.2658570, IEEE Transactions on Visualization and Computer Graphics

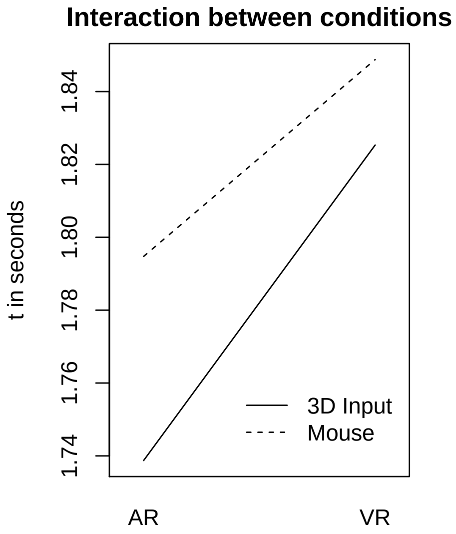


Fig. 6: Interaction effect between the conditions.

H_2: participants performed in fact ≈ 11.7 seconds or $\approx 14.8\%$ ($p < 0.040$ on a paired (within-subject) t-Test) faster in the AR condition.

This raises interesting questions as to what the observed speed-up in AR over VR can be attributed to. The ability to see ones' own hand was only relevant in the 3D Input Device condition, so a naive estimate could be that $\approx 40\%$ of the observed performance improvement in the 3D input device condition stems from the improved hand-eye coordination due to visual feedback. However, we have no means of validating this assumption. A mundane explanation for this difference in task performance which we cannot rule out is that participants simply preferred the different background. The VR environment was mostly white, except a base that resembled a circular checkerboard. The AR environment was in a gray pattern with faintly saturated color spots. Therefore, some participants may have found the virtual objects to be more clearly visible against the darker background, even though no participant mentioned anything similar.

In order to investigate the reasons behind the differences in performance, we performed three post-hoc analyses on additional data retrieved from the log files of the experiment. We analyzed the recorded translational motion data of the HMD (Figure 4(b)) and found that participants moved their head more when using the 3D input device than when using the mouse (on average $\approx 50\%$ increase; $F(1,489) \approx 200$, $p < 10^{-15}$ on a repeated-measures ANOVA). This is to be expected, since the increased degrees of freedom and range of motion of hand movements may cause secondary motion in the head. However, we also found that even in the mouse conditions, participants moved their heads significantly more in AR than in VR (on average $\approx 40\%$ increase; $F(1,489) \approx 50$, $p < 10^{-11}$ on a repeated-measures ANOVA). This raises the question whether participants in the AR condition were looking at the mouse. In order to verify this possibility we further analyzed the ratio between horizontal head rotation (yaw) and vertical head rotation (pitch) and found that on average, horizontal rotation was prevalent in all conditions. Interestingly, the dominance of horizontal rotation over vertical rotation was even more pronounced in the AR conditions ($\approx 19\%$ increase in both the mouse and 3D input device conditions) than in the VR conditions ($\approx 14\%$ increase when using the 3D input device, only $\approx 8\%$ when using the mouse), meaning that participants on average even reduced their relative vertical head rotation in the AR conditions in favor of more horizontal rotation. However, these findings were not statistically significant, indicating that the ratio between horizontal and vertical head rotation was less based on experimental conditions than on personal preference. Since the cameras attached to the HMD had a 75° vertical field of view and the mouse was positioned on a board on the participants' lap, we estimate that in order to observe the mouse, a 50° downward angle is required. However, only one single participant achieved such a low angle in only three measurements of one single condition (AR, using the 3D Input Device). The median lowest angle in all measurements was 18° (the mean lowest angle was 19.6° with a standard deviation of 10°). This strongly indicates that the increase in performance and vigor in the AR conditions was unrelated to the ability to look at the mouse. Another possibility is that participants were more effective simply because they were more engaged, without having a better strategy to solve the task. In order to verify this theory, we analyzed the recorded mouse motion, and again found a significant difference between AR and VR. In AR, participants on average moved the mouse more than 60% faster than in VR ($p < 10^{-13}$ on a within-subject t-Test; Figure 4(c)). This could either mean that the VR environment had a "stifling" (due to spatial unawareness) or "calming" (due to reduced visual load) psychological effect, or that the AR condition was just "more exciting" to our participants. It should be noted that since these three analyses were non-planned post-hoc comparisons, the required significance level per test was corrected to $\alpha \approx 0.01667$ (Bonferroni) or $\alpha \approx 0.01695$ (Šidák) in order to keep the Type I error rate at $\alpha = 0.05$ overall.

Regarding the subjective level of comfort, we found no significant difference between all four conditions. While it was obvious during the execution of the user study that most users found some conditions more comfortable than others, in the overall analysis these individual preferences did not produce a significant overarching tendency towards any of the conditions. A two-way repeated measured ANOVA showed only non-significant differences ($F(1,93) \approx 0.18$, $p \approx 0.67$ for mouse vs. 3D input device, $F(1,93) \approx 0.56$, $p \approx 0.46$ for AR vs. VR; Figure 7).

Therefore, H_3 was not supported, neither for differences regarding the input device and regarding the environment. One reason for this could be excitement over the novel input method combined with the rather short use time. Effects that are detrimental to comfort such as arm strain might not be felt immediately by participants who are excited about the ability to directly interact with virtual objects. Longer sessions might yield different results.

The same can be theorized about the missing personal preference between AR and VR. In longer work sessions, people might prefer the ability to see their surroundings in order to interact with real objects or communicate with their peers more naturally, but in our short task, no general trend became apparent.

4 CONCLUSION

In this study, we investigated the difference between AR and VR in an object selection and transformation task and found that AR consistently outperforms VR, even when using a 2D mouse as input device. While we found some indications as to why this might be the case, further research is required to determine specific factors and their effect.

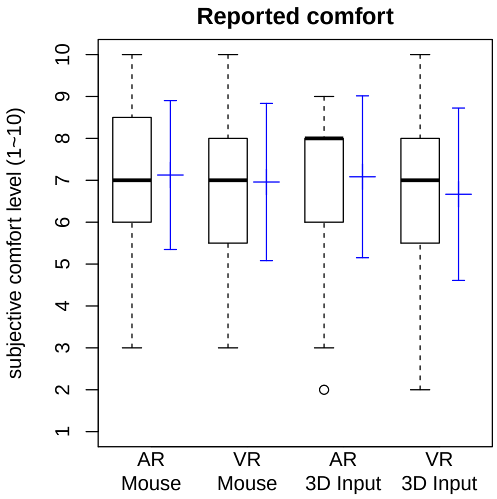


Fig. 7: Subjective level of comfort. Blue bars indicate mean and standard deviation.

One important question is how different physical environments affect users. Our test environment was a separated area of approximately $1.5m^2$ in one corner of the room, held in a neutral gray color and without the possibility of people getting close to the participant. Real-life conditions that differ from this set-up may produce different results. A wider area might encourage users to move more, even in a purely VR environment, or clutter in the background might distract users in the AR condition. People walking around in the area could also have an effect, as it may make users feel uneasy not being able to see them in a VR set-up. The possibility to rest one's elbows on a table surface may improve comfort and precision when using 3D input devices and may thus affect performance as well.

Another important factor in AR is the quality of the video images. The cameras that we used only provide a low frame-rate, low resolution, and suffer from visible delay. Better cameras may influence the results even more in the favor of using AR. When using optical see-through HMDs, different factors come into play, such as differences in accommodation depth between the virtual objects and the real environment.

We also did not consider long-term effects. One possibility is that users suffered from a temporary disorientation when entering the VR environment, which might have waned during longer use sessions.

Finally, we have to acknowledge that in normal applications VR can be realized much easier than AR and will usually be more responsive and less computationally expensive since the system does not have to wait for camera images and perform undistortion and rectification algorithms. In our prototype, the performed algorithms were exactly the same, with the only difference being that we discarded the video image in the VR condition. We did this in order to eliminate effects from differences in system performance. In real-life applications, however, this very difference may be of importance.

4.1 Future Work

This work represents a step towards understanding differences between AR and VR, and opens up several new venues for future research work.

The most interesting question that this study raised is the fact that even when using a 2D mouse, seeing the real environment has a positive effect on task performance, even though it adds no immediately relevant information for the task. This points towards the possibility that AR increases a users engagement, or that reduction of environmental complexity has a slowing effect on users. The former could be tested with a simulated AR system [13] that displays alternate versions of the real environment, on the hypothesis that any virtual environment is preferable to a completely empty environment, even if it does not represent reality. For the latter possibility, one factor might be the users' trust in knowing the boundaries of the workplace. A future study where the system displays a simple bounding volume could shed more light on this possibility. In our study, we only considered a 2D mouse and a 3D wand which was co-located to the real hand position, but not a desktop 3D input device. Such a device is a third alternative that — while allowing 3D interaction — is not co-located with the real hand. The differences in task performance between AR and VR conditions when using a desktop 3D input device could be close to that of a mouse — stressing the importance of seeing one's own arm — or to that of the 3D wand — indicating that AR somehow may provide better spatial understanding when performing 3D transformations.

Our study only gives an estimate of the effect of all visual stimuli in AR combined and can therefore not predict which specific stimuli are important to task performance. Especially visual feedback of the own arms and hands may be a critical factor. Future studies are required to dissect the measured performance difference which we presented in this paper into its main factors.

We also only considered the case of users sitting in a chair. Given that the work area was small enough to make walking unnecessary it is possible that the results may be similar when users are standing upright. However, further studies are required to validate this assumption.

Finally, we only analyzed participants' head motion as an indicator of the source of the performance difference. Future studies could also gather data on eye movement, dilation of pupil size (cognitive load), heart rate, and other indicators of subjective experience.

ACKNOWLEDGMENTS

The authors would like to thank the participants in the user study, as well as Aitor Rovira, Jan Gugenheimer, Jarkko Polvi, and Alexander Plopski for their helpful feedback.

REFERENCES

[1] A. Boud, D. J. Haniff, C. Baber, and S. Steiner, "Virtual reality and augmented reality as a training tool for assembly tasks," in *Information Visualization, 1999. Proceedings. 1999 IEEE International Conference on.* IEEE, 1999, pp. 32–36.

[2] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization.* ACM, 2008, pp. 9–14.

[3] M. A. Cidota, R. Clifford, S. G. Lukosch, and M. Billinghurst, "Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality," in *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on.* IEEE, 2016, pp. 172–177.

[4] M. C. Juan and D. Pérez, "Using augmented and virtual reality for the development of acrophobic scenarios. comparison of the levels of presence and anxiety," *Computers & Graphics*, vol. 34, no. 6, pp. 756–766, 2010.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2017.2658570, IEEE Transactions on Visualization and Computer Graphics

[5] J.-J. Arino, M.-C. Juan, J.-A. Gil-Gómez, and R. Mollá, "A comparative study using an autostereoscopic display with augmented and virtual reality," *Behaviour & Information Technology*, vol. 33, no. 6, pp. 646–655, 2014.

[6] S. M. Botden, S. N. Buzink, M. P. Schijven, and J. J. Jakimowicz, "Augmented versus virtual reality laparoscopic simulation: what is the difference?" *World journal of surgery*, vol. 31, no. 4, pp. 764–772, 2007.

[7] C. Sandor, S. Uchiyama, and H. Yamamoto, "Visuo-haptic systems: half-mirrors considered harmful," in *EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint*. IEEE, 2007, pp. 292–297.

[8] S. Irawati, S. Hong, J. Kim, and H. Ko, "3d edutainment environment: learning physics through vr/ar experiences," in *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*. ACM, 2008, pp. 21–24.

[9] P. Rhienmora, K. Gajananan, P. Haddawy, M. N. Dailey, and S. Suebnukarn, "Augmented reality haptics system for dental surgical skills training," in *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2010, pp. 97–98.

[10] C. Lee, G. A. Rincon, G. Meyer, T. Hollerer, and D. A. Bowman, "The effects of visual realism on search tasks in mixed reality simulation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 4, pp. 547–556, 2013.

[11] A. Möller, M. Kranz, S. Diewald, L. Roalter, R. Huitl, T. Stockinger, M. Koelle, and P. A. Lindemann, "Experimental evaluation of user interfaces for visual indoor navigation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 3607–3616.

[12] M. Khademi, H. M. Hondori, L. Dodakian, S. Cramer, and C. V. Lopes, "Comparing pick and place task in spatial augmented reality versus non-immersive virtual reality for rehabilitation setting," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 4613–4616.

[13] D. A. Bowman, C. Stinson, E. D. Ragan, S. Scerbo, T. Höllerer, C. Lee, R. P. McMahan, and R. Kopper, "Evaluating effectiveness in virtual environments with mr simulation," in *Interservice/Industry Training, Simulation, and Education Conference*, 2012.

[14] S. Howlett, R. Lee, and C. O'Sullivan, "A framework for comparing task performance in real and virtual scenes," in *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*. ACM, 2005, pp. 119–122.

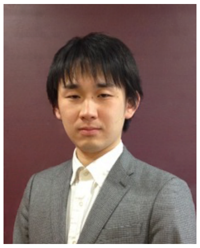
[15] P. J. Werkhoven and J. Groen, "Manipulation performance in interactive virtual environments," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 40, no. 3, pp. 432–442, 1998.

[16] Y. Coello, "Spatial context and visual perception for action." *Psicologica: International Journal of Methodology and Experimental Psychology*, vol. 26, no. 1, pp. 39–59, 2005.

[17] M. S. Graziano, D. F. Cooke, and C. S. Taylor, "Coding the location of the arm by sight," *Science*, vol. 290, no. 5497, pp. 1782–1786, 2000.

[18] O. N. Keene, "The log transformation is special," *Statistics in medicine*, vol. 14, no. 8, pp. 811–819, 1995.

Max Krichenbauer is a doctoral student at Nara Institute of Science and Technology, Graduate School of Information Science, Interactive Media Design Lab. He received his B.Sc. from the University of Applied Sciences Upper Austria in 2010 and his M.Sc. from Nara Institute of Science and Technology in 2014. His research focus is 3D user intefaces for media production.



Goshiro Yamamoto is a Senior Lecturer in the Kyoto University, Japan. He received the B.E., M.E., and Ph.D. in Engineering degrees from Osaka University, Japan in 2004, 2006, and 2009 respectively. Dr. Yamamoto worked at Okayama University and Nara Institute of Science and Technology. In 2016, he joined Kyoto University Hospital. His major interests are in Human-Computer Interaction and Projection-based Augmented Reality.

Takafumi Taketomi received his B.E. degree in National Institute for Academic Degrees and University Evaluation in 2006. He received his M.E. and Ph.D. degrees in information science from Nara Institute of Science and Technology in 2008 and 2011, respectively. He has been an assistant professor at Nara Institute of Science and Technology since 2011.

Christian Sandor is an Associate Professor at one of Japan's most prestigious research universities, Nara Institute of Science and Technology (NAIST), where he is co-directing the Interactive Media Design Lab. In 2005, he obtained a doctorate in Computer Science from the Munich University of Technology, Germany under the supervision of Prof. Gudrun Klinker and Prof. Steven Feiner. Before joining NAIST, he directed the Magic Vision Lab.

Hirokazu Kato is a professor in the Nara Institute of Science and Technology (NAIST), Japan. He received his B.E., M.E. and Dr. Eng. degrees from Osaka University, Japan in 1986, 1988 and 1996, respectively. Since 2007, he has been with the Graduate School of Information Science of NAIST. He received the Virtual Reality Technical Achievement Award from IEEE VGTC in 2009, and Lasting Impact Award at the 11th IEEE International Symposium on Mixed and Augmented Reality in 2012.