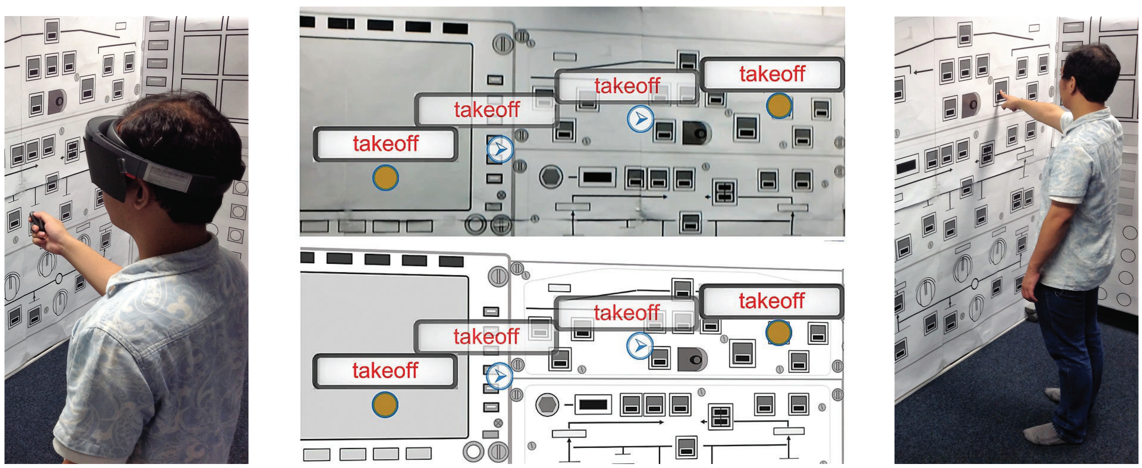
Transferability of Spatial Maps: Augmented Versus Virtual Reality Training

Nicko R. Caluya* Alexander Plopski Jayzon F. Ty
Christian Sandor Takafumi Taketomi Hirokazu Kato

Interactive Media Design Laboratory, Nara Institute of Science and Technology



(a) (b) (c)

Figure 1: An illustration of our experiment procedure showing the transfer of spatial maps. (a) The user wears the Microsoft Hololens HMD while performing a guided memorization task, where an icon moves along with a word label to a specific destination. (b) This task is done in two conditions: in AR (top) and in VR (bottom). (c) We confirm the memory transfer to the actual application in a memory transfer test 2 days after the training. The user was not wearing the HMD during this test.

Abstract

Work space simulations help trainees acquire skills necessary to perform their tasks efficiently without disrupting the workflow, forgetting important steps during a procedure, or the location of important information. This training can be conducted in Augmented and Virtual Reality (AR, VR) to enhance its effectiveness and speed. When the skills are transferred to the actual application, it is referred to as positive training transfer. However, thus far, it is unclear which training, AR or VR, achieves better results in terms of positive training transfer. We compare the effectiveness of AR and VR for spatial memory training in a control-room scenario, where users have to memorize the location of buttons and information displays in their surroundings. We conducted a within-subject study with 16 participants and evaluated the impact the training had on short-term and long-term memory. Results of our study show that VR outperformed AR when tested in the same medium after the training. In a memory transfer test conducted two days later AR outperformed VR. Our findings have implications on the design of future training scenarios and applications.

Index Terms: H.5.1—Information Interfaces and Presentation: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2—Information Interfaces and Presentation: Multimedia Information Systems—Ergonomics, Evaluation/methodology, Theory and methods—

*e-mail: {nicko.caluya.mv9, plopski, jayzon.ty.mv9, sandor, kato, takafumi-t} @is.naist.jp

1 Introduction

Worker training is a costly and time-consuming process. It involves learning a variety of steps that must be performed for each task, as well as memorizing important locations. For example workers in a control room must memorize all elements of the various control panels. In industry and manufacturing, the working environment often contains a large variety of objects placed in different locations within a large space. Being able to quickly locate objects in the environment is crucial for efficient task performance. As a result, various guidance systems have been developed in order to help users quickly locate specific objects and to build a spatial memory map of the environment [12, 14, 18].

Augmented and Virtual Reality (AR, VR) can assist the training process by presenting guidance to trainees, or simulating a variety of situations that could not be presented in reality [1, 2, 17].

In many cases, AR and VR could be applied for the same training scenario. For example, Dünser et al. [7] compared the effect of AR and a desktop CAD application on spatial ability training. They aimed to address improving the spatial ability of the users with four tasks (folding, cutting, rotation, orientation). Although they found that training in AR did not outperform the CAD application, it does not mean that AR and VR cannot assist in spatial memory training, which involves a larger area. It remains an open question which training environment would be better suited for spatial memory training.

In AR, the training is conducted in the target environment. One would expect AR training to have efficient skill transfer with a high retention rate from training to practical application. The main drawback of AR training systems is that the user must be present in the target environment. On the other hand, VR training allows users to train independent of their location, even from their homes. However, we expect VR training to have a lower retention rate of the information as the training conditions differ from the actual environment. It is unclear to what degree the learned content can be applied in practical situations, and how it differs from training in AR.

To better understand how well skills obtained in VR and AR are transferred to the actual application, we have designed a spatial memorization task in a control-room scenario (Fig. 1). We asked participants to memorize the location of elements associated with labels in AR or VR. In addition to a short-term memory test, we evaluated how well the learned material can be applied in a practical scenario via a memory transfer test conducted 2 days after the training. Our results show that although VR outperforms AR training in the immediate post-training test, it performs worse in the memory transfer test. The results of our study suggest that AR should be preferred over VR for spatial memory training. At the same time, we expect VR to perform similarly well if the training is repeated over an extended period of time.

The main contribution of our paper is being the first study to perform a comparison between the effectiveness of AR and VR in transferring a spatial map from a simulation to the actual work environment.

2 Related Work

In this section, we first give a brief introduction to the concept of spatial memory. We then look at studies that discuss how the environment design can assist spatial memory. Finally, we provide an overview of methods and systems that were developed for spatial memory training.

2.1 Spatial Memory

Spatial memory refers to the part of human cognition that makes us capable of retaining information about the geographic layout of an environment. Users with a spatial map of the environment can perform tasks more efficiently [3]. Scarr et al. [21] distinguish between two types of tasks associated with spatial memory: navigating through environments and remembering object locations. Spatial information can be learned either automatically or through effortful learning [13], i.e., some spatial information requires conscious effort from the user to retain it. Thus, there are two possible approaches to support spatial memory: present spatial information in a way that it supports automatic learning, or create methods or systems that encourage effortful learning.

2.2 Factors Affecting Automatic Spatial Learning

Various design factors can facilitate automatic learning of spatial maps. When we learn locations of objects in an environment, we tend to encode them relative to a spatial frame of reference [21], e.g., grids and landmarks. Leifert [16] investigated if grid-based structuring and grid lines have an effect on spatial and content memory. Her results show that while grid structuring has a positive influence on remembering the location of items, regardless of the presence of grid lines, it does not influence content memory. Furthermore, the presence of grid-lines yields worse performance on content memory. Uddin et al. [23] investigated if embedding artificial landmarks in grid-based user interfaces can positively influence spatial memory. They tested their hypotheses on grids with a small (64 items), medium (96 items), and large (160 items) number of items. Their results show that artificial landmarks improve spatial memorization in medium to large grids.

Fine and Minnery [9] investigated how saliency affects one's ability to remember target locations in a map. They found that the ability to recall locations of targets increases with target saliency, and that the effect becomes more apparent as the number of item locations increases. Santangelo and Emiliano [20] showed that users memorize locations of items better when they are placed over salient background.

Fujimoto et al. [10, 11] investigated how AR label placement affected user memory. They found that users are better at recalling labels placed in close proximity to the target than labels presented in a fixed location.

2.3 Effortful Memory Training

One of the oldest and most effective methods to enhance one's ability to recall information is the Method of Loci (MoL), also known as the "memory palace" technique [24]. Users memorize information by imagining themselves walking around a familiar environment, and placing items that they need to remember along the path. By walking along the same path and retrieving the items along the way, users can recall what they tried to memorize. While effective, MoL requires many hours of intensive training in order to work [15]. Even so, MoL remains one of the most widely used memorization techniques in present times. Legge et al. [15] asked users to traverse a path in a desktop virtual environment and to use it for MoL. They found that the virtual environment could be used as effectively as a familiar environment for MoL. Rosello et al. [19] created an AR version of MoL, where users could place objects into their surroundings to effectively create a memory palace. Their results show that using AR assisted users in creating a path and applying MoL. As a result, users could memorize items better than with conventional learning.

Cockburn et al. [4] designed a virtual keyboard interface that required users to brush off virtual "frost" in order to see the key label underneath. The frost re-appeared after a certain time. Users were thus encouraged to memorize the locations of the keys to avoid having to repeatedly remove the frost, which improved their spatial memory. Ehret [8] showed that spatial memory can also be enhanced through repeated searches for an item. Based on this observation, Kaufmann and Ahlström [14] investigated whether the use of a projector phone with a peephole interface can yield better spatial memory performance in a map navigation setting, as opposed to using a smartphone with a touchscreen interface. Their results showed that spatial memory performance is indeed better in the projector phone condition. Gacem et al. [12] investigated the impact of a projection-based visual guidance system on spatial memory, with consideration for type of control (user or system controlled) and kinesthetic feedback. They found that the system-controlled setup yields higher recall rates, and that kinesthetic feedback has no effect on recall rates, which contradicts previous findings, e.g., [22].

Although there have been studies that explored the benefits of VR and AR in spatial memory and task training, we are not aware of any work that compares the effectiveness of immersive VR and AR for spatial memory training. Our study aims to answer this question.

3 Experiment

The goal of our experiment was to compare the positive training transfer of AR and VR training. To prevent external effects on the training process, like muscle memory from walking to a specific lane in a warehouse, we chose the control room scenario, where participants have to memorize the function of a variety of control buttons and screens. This scenario is very similar to the one used by Gacem et al. [12]. In this section we describe the design of our experiment platform, the training procedure, the testing procedure, and the implementation of the training.

3.1 Experimental Platform

The results of the training can be affected by a variety of factors, such as the field of view (FoV) and resolution of the display, the realism of the scene, or the training scenario, to name a few. In our experimental platform, we tried to eliminate any external factors that could affect the results.

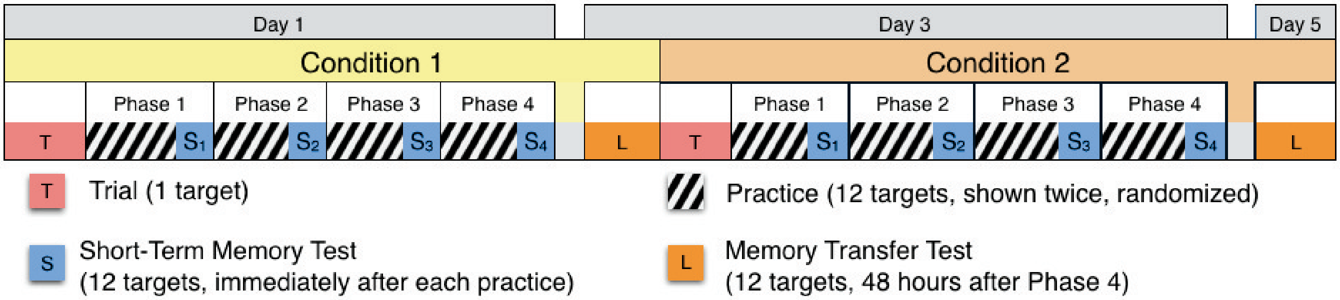


Figure 2: Experiment procedure mapping the two sessions over the course of five days.

3.1.1 Room Setup

In actual scenarios, control rooms have varying surface structures, and thus text legibility was also considered. DiDonato et al. [6] showed that users read texts better when it is in high contrast against flat surfaces. To help participants see the different targets on the wall, we printed the layout in grayscale on white background.

Our control room consists of 3 walls. The left and right walls measure 1.69m x 4m and the front wall measures 1.69m x 3.78m. On each wall, we placed posters with a variety of targets printed onto them. We also replicated this room and all targets as a virtual 3D model that was used in the VR training scenario.

Each wall of the control room was based on a generic design of a commercial airline cockpit interface. Each element of the setup was enlarged enough so that even the smallest components were visible from the center of the room. This presented us with a variety of elements, such as buttons, switches, and knobs, that were arranged in grids as well as standing out elements of varying sizes. Overall, the design had 17 unique elements with sizes ranging from 8x8 cm to 10x12 cm. Examples of the various elements are shown in Figures 1 and 4. We modified the layout to ensure that there was at least 4 cm spacing on all sides of a control room element, to prevent erroneous memorization due to minute misalignments of the CG. At the center of the frontal wall we placed an icon that represented a monitor. This area was used as an initialization area, where participants could see the label whose position they would learn next.

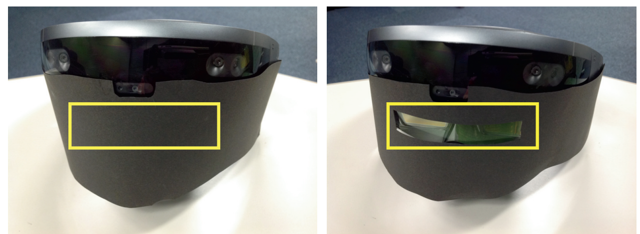
3.1.2 Hardware

The quality of AR and VR training can be affected by a variety of factors, such as FoV, resolution of the display, and image quality. To keep these factors constant, we used the same HMD for both training conditions. We used a Microsoft HoloLens, an optical-see-through head-mounted display (OST-HMD) that does not occlude the surrounding scene. For the VR condition, we used a black cover to block the participants' view of the real world environment, and to let them see only the virtual room (Fig. 3a). To ensure that participants had the same FoV in the AR and VR condition, we used a second cover that allowed participants to see only the area that would be augmented by the HMD in the AR condition (Fig. 3b). In both scenarios, the CG was rendered according to the positional tracking of the device.

3.1.3 Implementation

We developed this system using Unity 2017.1.0.f3 [1]. We replicated the training area as a virtual object that was enabled, thus visible to the user, during the VR training and short-term memory tests. The virtual model of the scene was disabled, thus invisible to the user, during the AR condition to allow a view of the training room with CG overlaid onto it.

[1]https://unity3d.com/



(a) VR (b) AR

Figure 3: Microsoft HoloLens masks for the two conditions. (a) In the VR condition the entire FoV is occluded by the mask, while (b) in the AR condition the area augmented by the CG (inside the yellow box) remains open.

3.2 Subjects

We recruited 16 participants from our university (13 male, 3 female, age 22 to 33 (25.75 ± 3.044)). All participants received compensation for their time with an amount equivalent to 10USD. Before taking part in the experiment, participants signed a consent form and answered a preliminary survey about their familiarity with AR, VR, as well as the Microsoft HoloLens. Results show that 9 participants (56.25%) had prior experience with AR, 10 participants (62.5%) had prior experience with VR, and 5 participants (32.25%) have used a Microsoft HoloLens in the past.

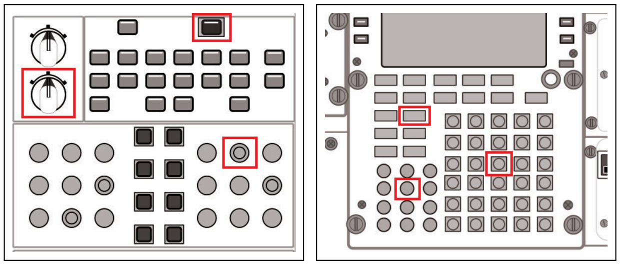
3.3 Procedure

We instructed the participants to follow certain steps to complete the experiment. Here we present an overview of the experiment procedure, followed by a detailed explanation of each component.

3.3.1 Overview

The goal of our study was to evaluate the quality of memory transfer. We conducted a within-subjects experiment over the course of 5 days. Figure 2 illustrates the structure of the experiment timeline. On the first day, participants would complete a trial session, denoted by T in Fig. 2, that consisted of learning the position of a single label to familiarize themselves with the system. They then took part in 4 training phases, with each phase consisting of a practice session followed by a short-term memory test (denoted by S_i in the figure, with i being the phase number, e.g., the last short-term memory test in the sequence is denoted by S_4).

On the third day, participants came back to take a memory transfer test, denoted by L in Fig. 2. After they completed the test, they started the second training session in the untrained condition. Finally, on the fifth day, participants came back for a memory transfer test of the second condition and provided feedback via free-form comments.



(a) Salient (b) Less Salient

Figure 4: Examples of salient and salient targets.

(a) Correct (Green) (b) Wrong (Red)

Figure 5: System feedback after the user selects a location in the short-term memory test.

During the training we asked participants to stand in the center of the room to allow them to see all walls. However we did not prevent them from taking some steps to the side, as this would be natural behavior that is possible in both scenarios. Furthermore, we did not require participants to perform any motion for kinesthetic memorization.

We selected 24 evenly distributed targets on the walls in terms of location and saliency. Hereby, saliency is determined by the uniqueness of an object (see Fig. 4). For example, we expect an object that is part of a grid pattern to be less salient than an object that is standing out distinctively. Similarly, a smaller object is less salient than a similar larger object. Each target was assigned a unique label based on terms related to aviation. The targets were divided into two datasets, wherein participants learned one dataset in the VR, and the other in the AR condition.

During each training session participants would see each label twice and the order in which the labels appeared was randomized. During the testing phase we checked all 12 targets covered in the training session in randomized order. For each part of the experiment we recorded the time participants took to complete the task, the number of correctly answered questions, as well as the errors.

3.3.2 Training Procedure

Our learning procedure consisted of three steps:

1. The target icon appears in the central area of the front wall.

2. After the user focuses on the target icon, it starts moving towards the goal location. Whenever the target would leave the user's view, it would stop.

3. After reaching its destination target, the icon changes color and disappears after 5 seconds, or after the user clicked a button on a controller.

The target icon consisted of a label and an icon that would show the direction towards the corresponding element on the wall. The label and icon were colored to help participants see them against the wall. In an ideal scenario, with a wide field-of-view HMD it would be possible to simply present the labels at the corresponding location and present only minimal guidance to help users build a spatial map. However, due to the limited field-of-view of the device used in our experiment we opted for presenting the label at an initial location and asked participants to follow it towards the corresponding location as in [12]. Contrary to [12], we moved the icon and the label together to the target location. This was based on the observation that users memorized items better, if the label was close by [10].

3.3.3 Testing Procedure

To verify the results of the training, we conducted a short-term memory and a memory transfer test. During these tests, participants had to point out the location of the elements that corresponded to labels they had learned during the training. The short-term memory test was conducted in AR/VR. Here participants had to align an icon with the corresponding element. After they confirmed their selection, the icon changed to show them whether their selection was correct (Fig. 5).

On the other hand, the training transfer test was conducted in the physical environment (regardless of the training condition). For each label, participants were asked to point out its location on the wall and we did not provide any feedback about the correctness of the result. Therefore, participants would not rely on the familiar training device and environment in order to apply the learned skill.

The choices for the number of targets per practice, as well as frequency and timing of the tests were based on the experiment conducted by [12]. These choices were also consistent with findings by [5], where the definitions of short-term memory and long-term memory were investigated in a thorough review of related literature in terms of temporal decay (duration) and ability to store chunks of information (capacity). Phases 1 to 4 containing the practices and short-term memory tests have no breaks, so as to remove effects of interferences. This makes 48 hours an adequate gap between the end of Phase 4 and the start of the memory transfer test. Finally, we chose 12 targets, which is more than the average 7 chunks that people can accommodate. We expect that participants are unlikely to achieve a 100% recall rate on S_1.

3.4 Independent Variables

We counterbalanced our experiment in terms of the order in which participants trained each condition (AR or VR first), as well as what dataset was used to ensure that this did not affect the results. Our study had the following independent variables:

Condition: Denotes whether participants were training in AR or VR.

Proximity: Denotes how close the target was with respect to the initialization area. We define a target as near to the starting area if it is on the front wall or is no more than 1.33m away from it (corresponding to the third of the side walls that is closest to the front wall). The rest of the side walls, was considered as far.

Saliency: Denotes if a target was unique or large and thus *salient*. On the other hand, targets that were part of a grid and smaller were *less salient* (Fig. 4).

Session: Denotes when the test scores were collected. In particular, we used the four training tests S_1, S_2, S_3 and S_4, and the memory transfer test L.

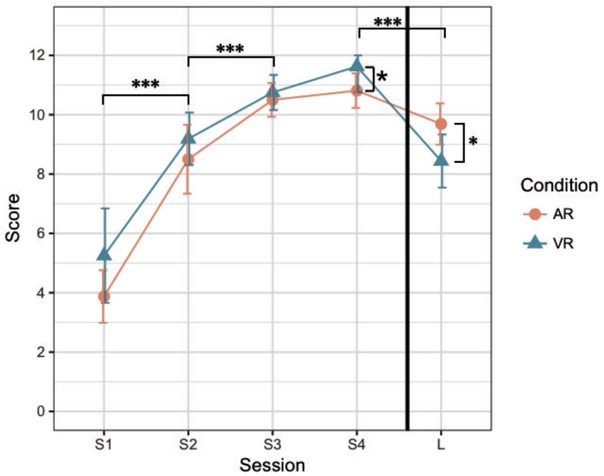


Figure 6: Mean scores per condition across all short-term tests (S_1-S_4) and the memory transfer test (L). The whiskers denote confidence interval. Connected bars represent significant difference (* = p<0.05, ** = p<0.01, *** = p<0.001).

3.5 Hypotheses

While the experiment attempts to investigate the difference between VR and AR conditions, we designed the training scenario to be as consistent and controlled as possible for both environments. We tried to ensure that the 3D model of the room matches the actual room in terms of layout and positions of all elements. With this in mind, we expect participants to retain information equally well during the training for AR and VR (**H1**).

We also expect that training in the actual environment will facilitate memory transfer. The AR condition lets the participants view the actual environment during training, compared to the 3D rendering of the walls occluding the actual environment in the VR condition. It would make sense that training transfer is easier in AR because the extent of the world knowledge is the same (i.e., the viewed environment during training is the same as when the actual task is performed). Thus we expect participants to recall labels they learned in AR better than those learned in VR during the long-term test (**H2**).

Furthermore, based on previous findings [10] we expect that the distance from the initialization location to the target location will not affect how well participants memorize the location because the labels travel along with the icon towards the target location (**H3**).

Finally, we expect that the saliency of the targets will affect how well the participants memorize them. (**H4**). Based on our definition of saliency, targets with a bigger size would catch more attention, thus, we speculate participants would pick up the location information much faster. We also expect that elements that we designed to stand out over a structured grid would emulate the artificial landmark observation by [23], which should yield better recall rates for these targets.

Table 1: Average recall percentages according to proximity.

Proximity	**Short Term**				**Memory Transfer**
	S1	*S2*	*S3*	*S4*	
Near	36.89%	76.89%	88.00%	94.22%	73.78%
Far	38.36%	69.81%	89.94%	93.08%	77.36%

4 Results

Both VR and AR sessions had similar training and testing durations. On average, training lasted for 21.94 minutes, and the memory transfer test took 4.06 minutes in the VR condition. Meanwhile, in AR the training sessions lasted for 19.31 minutes, and the memory transfer tests for 4.75 minutes. A paired t-test reveals that there is no significant difference between the two conditions (Training: t = -1.324, p = 0.205; Memory transfer test: t = 0.983, p = 0.34).

Aside from time, we also recorded the number of correct responses of the participants on all tests. We show the recall rates for each test in Fig. 6.

We analyzed the training data using repeated measures ANOVA with *Condition* × *Saliency* × *Proximity* × *Session*, using all 4 training sessions. The results show that *Session* ($F_{3,45} = 102.320, p < 0.0005$) and *Saliency* ($F_{1,15} = 15.882, p < 0.001$) had a statistically significant impact on the results during the short-term training tests. We did not find any statistically significant interaction between the different factors.

. We also compared the performance between the individual tests. One-way ANOVA test revealed a statistically significant difference between S_1 and S_2 ($F_{1,15} = 96.71, p < 0.001$), S_2 and S_3 ($F_{1,15} = 43.55, p < 0.001$), and S_4 and the memory transfer test L ($F_{1,15} = 22.16, p < 0.001$).

To better understand how *condition* affected the results at each training stage, we performed one-way ANOVA tests for each session. We found a statistically significant difference between the AR and VR conditions in the scores for S_4 ($F_{1,15} = 4.88, p = 0.04$). We did not find any statistically significant difference between participants' ability to recall *salient* targets compared to *less salient* targets for S_4.

We evaluated the results of the memory transfer test with a repeated measures ANOVA *Condition* × *Saliency* × *Proximity*. The results show that *Condition* had a statistically significant effect on the results ($F_{1,15} = 6.229, p = 0.025$). We also found statistically significant interaction between *Saliency*, *Condition*, and *Proximity* ($F_{1,15} = 7.3, p = 0.016$). A post-hoc Tukey Test showed that there was a statistically significant difference between users' recollection of far salient targets that were trained in AR and near less salient targets that were trained in AR.

We also checked whether prior experience with the device had any effect on participants' performance, regardless of AR or VR condition. We compared the mean scores (short-term and memory transfer) between participants with and without prior experience with a t-test. Although the sample size of our experiment is too small to be generalizable, the results reveal no significant difference on both the last short-term memory test S_4 (VR: $t = 0.13, p = 0.90$; AR: $t = -0.75, p = 0.47$) and L (VR: $t = 0.04, p = 0.97$; AR: $t = -0.82, p = 0.43$).

In this experiment, we have prepared two sets of target locations. We investigated if the two datasets were comparable in difficulty with one-way ANOVA and found no statistical difference between the results of the two datasets ($F_{1,15} = 2.00, p = 0.18$). The order in which participants trained in the conditions also did not have a statistically significant impact on the results ($F_{1,15} = 1.15, p = 0.3$).

We also investigated how *proximity* of the targets to the initialization area affected their recall rate (Table 1). Repeated measures ANOVA did not reveal any statistically significant interaction between *session* and *proximity*.

Finally, we also investigated how *saliency* of the targets affected participants' recall rates (Table 2). Repeated measures ANOVA did not reveal any statistically significant interaction between *saliency* and *condition*.

5 Discussion

As expected, the participants learned the label locations over the course of multiple training sessions, as shown by significant differences between consecutive short-term test scores. The scores plateaued at the end, resulting in no statistically significant difference between the third and last short-term test scores.

We believe that the test results can be explained by how well the training matched the application scenario. While the environment model used in the VR training matched the real scene, it was rendered with simple shading and no illumination. This resulted in low user cognitive loads and thus more efficient learning. On the other hand, the environment that the user was presented with in the AR training scenario included shadows, flickering light, and potentially transparent CG. Therefore, it better matched the actual application scenario, whilst increasing the users' cognitive load during training.

The results of the last short-term test reject our hypothesis **H1**. From Figure 6 we can see that VR training consistently outperformed in the short-term tests. One aspect could be that users were less distracted from the task by occluding the environment thus excluding any potential distractors, like flickering lights, shadows, etc. that may have distracted users in AR. Another reason could be that the participant's cognitive workload in the VR condition was lower than in the AR condition, as the visual information only presented the rendered CG without lighting and shading factors. This simplified training environment may have also contributed to easier and faster learning, and better performance in the short-term tests.

On the other hand, users showed a sharp drop in the recall rate when moving from the VR to the real environment during the memory transfer test. Although the performance for the AR condition decreased as well, the recall rate was significantly higher than for the VR condition. This supports our hypothesis **H2**. Most likely, the association of the training and the testing environment seems to have been stronger in the AR condition, which led to better memory transfer. The AR condition may have induced a more effortful training, such that there was better retention of information. This also implies that training in VR must be complemented with practical sessions to assist the memory transfer.

Although we did not find any statistically significant effect of the training condition on the recall rate of salient and less salient features during the memory transfer test, plotting the corresponding recall rates showed that while salient features were recalled equally well for both conditions, participants recalled less salient features less frequently after the VR training. In the future, further experiments with a larger number of participants, and more complicated tasks are necessary to determine if there is a potentially significant effect.

Our results contradict the findings of Gacem et al. [12]. This is likely because in their study the label always remained on a single position, while in our case the label moved towards the location. This helped users associate the label with the location [10]. However, even if the results support our hypothesis **H3**, there might still be subtle differences among participants when memorizing near or far targets.

Finally, the results reject our hypothesis **H4**. This may be attributed to how we defined what is salient and what is less salient. Other properties of a target, like the saliency of its outline against the white background, could have assisted the memorization. Complexity of the structured grid design can also be looked at (e.g. number of elements, variations of elements in the structure). While we have expected that participants will be more likely to make mistakes on targets located in a grid, some participants reported that they memorized these locations like an entry of a matrix (i.e. putting the label, row number, and column number in one chunk). Thus, these participants gave much more effort and attention in memorizing less salient targets which belong to a structured grid. It is also possible that the targets we deemed as small were large enough to be easily memorized by the participants.

In the next section we report some of the observations we made during the experiment and evaluation of the results.

Table 2: Average recall percentages according to saliency of target.

Saliency	**Short Term**				**Memory Transfer**
	S1	*S2*	*S3*	*S4*	
Salient	43.17%	77.97%	92.07%	94.71%	79.30%
Less Salient	29.94%	67.52%	83.44%	91.72%	70.70%

5.1 Error Analysis

As participants made fewer and fewer errors as the training progressed we observed what kind of errors were made throughout the training. Out of 1,920 trials across all tests (short-term and memory transfer), there were 512 errors. For the first test, the majority of the errors was incorrect assignment of labels to targets of the same dataset. For the second dataset, we additionally observed that participants selected targets that were part of the dataset they had learned in the other condition. This is in line with the observations made by Leifert [16], who reported that users are better at recognizing locations than the content associated with them. Adjacency of targets also had an effect on memorization. In our setup we had two big targets that were next to each other and belonged to the same grid, however were assigned to different datasets. One participant switched the answers and two participants mentioned that they paid special attention to these targets because of the adjacency.

5.2 Observations

In our experiment we ensured that the virtual room model is a close reconstruction of the actual room, this is supported by some participants mentioning that they did not notice the difference between the two conditions. At the same time, other participants noted that the obvious materials used in creating the actual environment (e.g. printed paper corners, tape) somehow served as landmarks [23]. These material details do not appear in the VR condition, as the 3D model of the wall occluded these features. Since the level of realism may vary from one training scenario to another, designers must exercise caution when dealing with the issue of realism of the virtual objects and the virtual world. These observations further raise the questions how the realism and truthful reconstruction of the scene affect the results as well. Given the results of our study, one would expect that an imperfect reconstruction would lead to even lower retention rates. However, this remains to be verified in the future.

In both conditions, some participants moved one or two steps back in order to see a increase the area they could see through the display. This suggests that when training on HMDs, the field of view could have an impact on how well users memorize the location of targets, and we plan to investigate this in the future.

6 Conclusions

In this paper we compared how spatial memory from AR and VR training transfers to actual applications. Although we found that VR performs better in the immediate post-training tests, it was performing significantly worse in the memory transfer test. This suggests that AR is better suitable for spatial memory training, especially if users cannot perform multiple training sessions. Nonetheless, VR remains a viable tool for spatial memory training, as it is easier to repeat the training in VR, compared to AR.

With the spread of affordable head-mounted displays on-the-job training of users through AR will become even easier and more appealing, especially if it does not interfere with the overall workflow. It is also preferred to training in VR as it helps retain the learned information and transfer it to the actual scenario. Nonetheless, VR can compensate the training to further reduce the overall training and memorization time.

7 Limitations and Future Work

There are several limitations to our study, which we would like to explore in the future. One such limitation is the fact that we used a Microsoft HoloLens to simulate an immersive VR experience. The small FoV of the HoloLens is very different from the typical immersive VR HMDs that have a large field of view. As such, our test was not a perfect simulation of an immersive VR experience. This fact coupled with the observation that some participants took a couple steps back to better observe the scene raises the question how the FoV of the display affects memory transfer. The generalizability of our study is thus limited by the hardware and the simple scenario that we used. In the future, we want to investigate how the FoV of the HMD affects the recall rate and if the results of our study can be replicated with wide FoV OST-HMDs.We also plan to investigate how other factors, like environment setup affect the memory, and if it could be used in a similar way to an MoL.

Another limitation is that our study did not involve any steps that would involve muscle memory, like walking, or handling of tools. In the future, we also want to compare how performing tasks in AR and VR affects the user's performance, and whether skills gained from training in AR/VR systems can be transferred to the actual workspace, i.e., training transfer.

Acknowledgments

We would like to thank the participants in the study for their time. We would also like to thank the reviewers for their valuable comments and suggestions. This research is funded in part by Grant-in-Aid for Young Scientists (B), #17K12726 from Japan Society for the Promotion of Science (JSPS), Japan.

References

[1] J. Bliss, P. Tidwell, and M. Guest. The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters. *Presence: Teleoperators and Virtual Environments*, 6(1):73–86, 1997.

[2] S. Büttner, H. Mucha, M. Funk, T. Kosch, M. Aehnelt, S. Robert, and C. Röcker. The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach. In *Proceedings of the International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 433–440, 2017.

[3] A. Cockburn, C. Gutwin, and S. Greenberg. A Predicitve Model of Menu Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computer Systems*, pp. 627–636, 2007.

[4] A. Cockburn, P. O. Kristensson, J. Alexander, and S. Zhai. Hard Lessons: Effort-inducing Interfaces Benefit Spatial Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1571–1580, 2007.

[5] N. Cowan. What are the Differences Between Long-term, Short-term, and Working Memory? *Progress in Brain Research*, 169:323–338, 2008.

[6] M. D. Donato, M. Fiorentino, A. E. Uva, M. Gattullo, and G. Monno. Text Legibility for Projected Augmented Reality on Industrial Workbenches. *Computers in Industry*, 70:70 – 78, 2015.

[7] A. Dünser, K. Steinbügl, H. Kaufmann, and J. Glück. Virtual and Augmented Reality as Spatial Ability Training Tools. In *Proceedings of the ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Design Centered HCI*, pp. 125–132, 2006.

[8] B. Ehret. Learning Where to Look: Location Learning in Graphical User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 211–218, 2002.

[9] M. Fine and B. Minnery. Visual Salience Affects Performance in a Working Memory Task. *Journal of Neuroscience*, 29(25):8016–8021, 2009.

[10] Y. Fujimoto, G. Yamamoto, H. Kato, and J. Miyazaki. Relation Between Location of Information Displayed by Augmented Reality and User's Memorization. In *Proceedings of the Augmented Human International Conference*, pp. 7:1–7:8, 2012.

[11] Y. Fujimoto, G. Yamamoto, T. Taketomi, J. Miyazaki, and H. Kato. Relationship Between Features of Augmented Reality and User Memorization. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pp. 279–280, 2012.

[12] H. Gacem, G. Bailly, J. Eagan, and E. Lecolinet. Impact of Motorized Projection Guidance on Spatial Memory. In *Proceedings of the ACM Symposium on Spatial User Interaction*, pp. 51–59, 2016.

[13] L. Hasher and R. Zacks. Automatic and Effortful Processes in Memory. *Journal of Experimental Psychology: General*, 108(3):356–388, 1979.

[14] B. Kaufmann and D. Ahlström. Studying Spatial Memory and Map Navigation Performance on Projector Phones with Peephole Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3173–3176, 2013.

[15] E. Legge, C. Madan, E. Ng, and J. Caplan. Building a Memory Palace in Minutes: Equivalent Memory Performance using Virtual versus Conventional Environments with the Method of Loci. *Acta Psychologica*, 141(3):380–390, 2012.

[16] S. Leifert. The Influence of Grids on Spatial and Content Memory. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 941–946, 2011.

[17] U. Neumann and A. Majoros. Cognitive, Performance, and Systems Issues for Augmented Reality Applications in Manufacturing and Maintenance. In *Proceedings of the International Symposium on Virtual Reality*, pp. 4–11, 1998.

[18] R. Reif and W. Gunthner. Pick-by-Vision: An Augmented Reality Supported Picking System. In *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pp. 57–64, 2009.

[19] O. Rosello, M. Exposito, and P. Maes. NeverMind: Using Augmented Reality for Memorization. In *Proceedings of the Annual Symposium on User Interface Software and Technology*, pp. 215–216, 2016.

[20] V. Santangelo and M. Emiliano. Visual Salience Improves Spatial Working Memory via Enhanced Parieto-Temporal Functional Connectivity. *Journal of Neuroscience*, 33(9):4110–4117, 2013.

[21] J. Scarr, A. Cockburn, and C. Gutwin. Supporting and Exploiting Spatial Memory in User Interfaces. *Foundations and Trends in Human–Computer Interaction*, 6(1):1–84, 2013.

[22] J. F. Soechting and M. Flanders. Sensorimotor Representations for Pointing to Targets in Three-Dimensional Space. *Journal of Neurophysiology*, 62(2):582–594, 1989.

[23] M. S. Uddin, C. Gutwin, and A. Cockburn. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3843–3855, 2017.

[24] F. A. Yates. *The Art of Memory*. Random House UK, 1966.