Erschienen in: CHI'16 : Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems /
Kaye, Jofish et al. (Hrsg.). - New York, NY : ACM Press, 2016. - S. 1245-1249. - ISBN 978-1-4503-3362-7
https://dx.doi.org/10.1145/2858036.2858043

Virtual Objects as Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load

Jens Müller, Roman Rädle, Harald Reiterer
Human-Computer Interaction Group, University of Konstanz
{Jens.Mueller,Roman.Raedle,Harald.Reiterer}@uni-konstanz.de

ABSTRACT

In collaborative activities, collaborators can use physical objects in their shared environment as spatial cues to guide each other's attention. Collaborative mixed reality environments (MREs) include both physical and virtual objects. To study how virtual objects influence collaboration and whether they are used as spatial cues, we conducted a controlled lab experiment with 16 dyads. Results of our study show that collaborators favored the virtual objects as spatial cues over the physical environment and the physical objects: Collaborators used significantly less deictic gestures in favor of more disambiguous verbal references and a decreased subjective workload when virtual objects were present. This suggests adding additional virtual objects as spatial cues to MREs to improve user experience during collaborative mixed reality tasks.

Author Keywords

Mixed reality; collaboration; virtual spatial cues.

INTRODUCTION

Mixed reality (MR), as introduced by Milgram and Kishino [11], describes the blending of physical and virtual objects on a single display. Virtual objects are rendered on top of a video see-through display which creates the illusion as if they were situated in the same physical space (see Figure 1). Users can benefit from such MR applications when viewing and manipulating virtual objects becomes a familiar physical interaction. MR has been proven not only to be beneficial for single user applications such as education, manufacturing, and architecture [2], but has also been proposed as a tool for computer supported collaborative work. Billinghurst [3], for instance, refers to two inherent qualities of MREs that are crucial to collaboration: First, MREs provide seamless transitions between the shared workspace (the task area in which collaborators are situated) and the speakers' interpersonal space (the communication space which allows for social interactions). Second, MR can enhance reality and may thereby "satisfy the needs of communication." [3] Closely related to the question of how to enhance reality in MR is the aspect of artificiality, which can be described as "the extent to which a space is either synthetic or is based on the physical world". [1]

Figure 1. Dyads solving an object identification task in a mixed reality environment with additive virtual objects (e.g., armchairs and a vending machine) that serve as spatial cues.

For collaboration, shared visual information such as spatial cues are known to play a crucial role to coordinate collaborators' actions (see [5–7,13]). Given that virtual objects are highly customizable in their appearance and their behavior (e.g., they do not represent physical obstacles), it raises the question whether spatial cues can be synthesized in MREs. In this note we investigate how additive virtual objects shape communication behavior and user task load in co-located MREs.

RELATED WORK

Our work is based on two strains of research: *influence of visual cues on individual cognition* and *influence of visual cues on group coordination*. For each strain, we introduce related work and summarize with the formulation of a hypothesis.

Influence of Visual Cues on Individual Cognition

The importance of visual information for cognition during navigation tasks has been well established both in physical and in virtual environments. Montello [12] discusses the visual and structural properties of physical environments and how they determine the ease of navigation therein. Based on research on landmark design in physical environments, Vinson [14] provides a set of design guidelines to support navigation in virtual environments (VRs). A reoccurring aspect refers to the "distinctiveness" of landmarks, e.g., that "a landmark must be easy to distinguish from nearby objects and other landmarks," and that "Landmarks must carry a common element to distinguish them, as a group, from data objects." [14]

Influence of Visual Cues on Group Coordination

The importance of visual information in a shared workspace has been established in terms of the coordination of collaborative activities (e.g. [5–7,13]). In the field of MR, Kiyokawa et al. [8] investigated communication behaviors in co-located collaborative Augmented Reality interfaces. In a target identification task they found that "the more difficult it was to use non-verbal communication cues, the more people resorted to speech cues to compensate." [8] Chastine et al. [4] investigated referencing behaviors in a collaborative modelling task in an MRE. They found that groups make heavy use of deictic speech, and that with increasing concreteness of the model, some groups created their own vocabulary to identify elements of the workspace.

These findings show that 1) spatial cues play a crucial role for individual cognition during navigation tasks, 2) that they represent an important coordination mechanism in collaborative tasks, and 3) that they can be synthesized in VRs. Based on these findings, we hypothesize that spatial cues can also be synthesized in MREs and that their presence positively influences collaboration.

EXPERIMENT

To investigate how additive virtual objects influence collaboration in co-located MREs, we conducted a controlled lab experiment. The study used a counterbalanced within-subjects design with the provision of additive virtual objects being the independent variable ("additive" in the sense of "additional to the actual data objects the participants had to work with," henceforth referred to as "objects" and "no objects"). The dependent variables were user task load (NASA TLX), communication behavior (video analysis), user experiences (semi-structured interview) and spatial memory" (reconstruction task).

Participants

We recruited 32 participants (8 female, 24 male) between 16-39 years of age (*M* 26.06, *SD* 5.63), forming 16 dyads. 19 participants were university students, 9 employed persons, and 4 secondary schools students. 4 participants reported prior experiences with MR technologies. 16 participants indicated regular tablet usage.

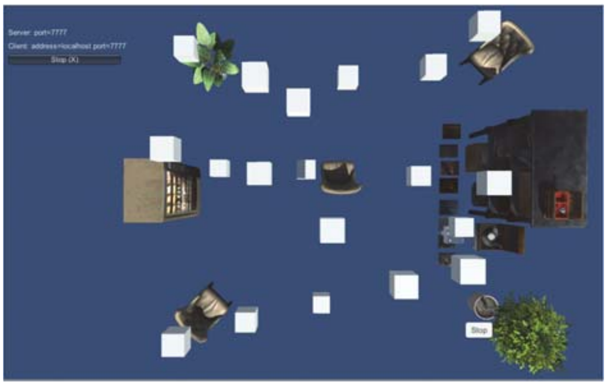

Figure 2. Bird's eye view of the virtual MR space in the condition with additive objects (including the memory cubes).

Apparatus and Study Environment

In our lab we allocated physical space of 4×4×2m, where participants could move freely. As MR displays we used Project Tango Tablets (370g, 1920x1200 pixels on a 7.02″ display (323 ppi) [15]). Due to the tablets' capability of area learning no additional tracking hardware was required to locate their position and orientation in space (6DOF). This guaranteed a high external validity since they can also be used outside of research lab facilities. We developed an MRE with Unity [16]. As potential physical cues we placed a waste paper basket, a clothes hook, a chair, a double ladder, several wall papers, and a floor lamp at the border of the interaction space. As virtual objects we used three armchairs, a bookshelf, two house plants and a vending machine (see Figure 2).

Study Task

For the design of our study task we referred to spatial planning tasks (such as architecture [3]). Such tasks require the collaborators to identify particular objects in the workspace (henceforth *object identification task*). There are also situations, when collaborators need to position a virtual object at a specific position in the workspace (henceforth *object positioning task).*

Object identification task

To create a dynamic situation in which collaborators have to exchange spatial information, we let collaborators play a modified (3D) version of the memory card game. In our version there were 10 pairs of white memory cubes as data objects (25cm edge length) which were randomly distributed in the MRE (see Figure 2). Each pair was texturized with the same symbol from the Wingdings font. Cubes were initially in the "covered" state. A cube could be "uncovered" by touching it. Unlike the original version of the game, dyads had to find matches collaboratively. Due to the collaborative nature of our version, collaborators benefited from each other's spatial knowledge, whereby communication was stimulated. In line with the original memory game, non-matching pairs of uncovered cubes had to be covered to continue with the next move. Accordingly, matches were removed from the MRE.

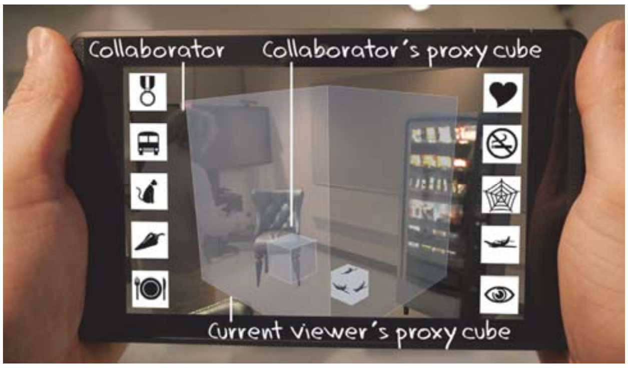

Figure 3. A semi-translucent proxy cube (displayed in front of each tablet) provides a dynamic positioning hint.

Object positioning task

The object identification task was followed by an object positioning task. In this task, dyads had to collaboratively position the memory cubes within the MRE according to their positions in the preceding objects identification task. The 10 symbols were displayed as buttons on each tablet (see Figure 3). A virtual semi-translucent proxy cube was displayed 0.5m in front of each tablet, which allowed participants to estimate the position where a memory cube would appear in the MRE when the corresponding button was pressed. To enable collaborative fine-tuning of positions, participants could see each other's proxy cube on their display. Deposited cubes could be repositioned.

Procedure

Participants were welcomed and introduced to the study. Afterwards, they were asked to fill out a demographic questionnaire. Then participants were introduced to the object identification task and were provided with a training phase (no additive virtual objects were provided, and a test set of symbols and coordinates was used for the memory cubes) to familiarize themselves with the devices and the task. Then, the objects identification task started in their assigned first study condition (≈ 10 mins). Afterwards, the NASA TLX questionnaire was provided to the participants. Then, participants started with the object positioning task in the same condition (≈ 15 mins). Afterwards, participants were provided with the NASA TLX questionnaire. This procedure was repeated in the respective other condition. After completion of the tasks in the second study condition a concluding, semi-structured interview on participants' experiences was conducted. Each session took about 60 minutes. Participants were compensated for their time.

RESULTS

The reporting of study results is structured into the two study tasks. Non-parametric tests were used when the assumption of normal distribution was violated. Results are marked with the subscript $_{NO}$ for the condition when no additive virtual objects were provided, and $_{O}$ when additive virtual objects were provided.

Object identification task

For analysis of *communication behavior,* video material from half of the sessions was analyzed for participants' spatial expressions. A cluster analysis yielded a set of 8 categories (see Figure 4): *deictic speech* (spatial indications, which "cannot be fully understood by speech alone" as explained by Kiyokawa et al. [8], e.g., "here," "over there,"), *participant* (e.g., "in front of me"), *region* (e.g., "in the center of the room"), *physical object* (e.g., "at the chair"), *other virtual cube* ("left of the uncovered cube"), and the *virtual object* ("near the shelf"). For non-verbal spatial expressions we identified *hand gestures* as the most prevalent ones. *"Non-assigned"* summarizes spatial expressions that occurred extremely seldom (e.g., pointing on the tablet, head-, and feet gestures) and those that could not be assigned to one of the other categories. Finally, all videos were coded using these categories as a coding scheme. A Wilcoxon signed rank test revealed that deictic speech, virtual cubes, and hand gestures were used significantly less frequently when objects were provided (deictic speech: $M_{NO} = 18.1$, $SD_{NO} = 8.5$, $M_{O} = 10.3$, $SD_{O} = 4.5$, $p = .010$; virtual cube: $M_{NO} = 4.1$, $SD_{NO} = 3.0$, $M_{O} = 2.2$, $SD_{O} = 1.4$, $p = .025$; hand gesture: $M_{NO} = 3.8$, $SD_{NO} = 3.6$, $M_{O} = 1.9$, $SD_{O} = 2.5$, $p = .039$) (see Figure 4).

User task load was analyzed with the Wilcoxon signed rank test. The test revealed a significantly lower cognitive task load when virtual objects were provided ($M_{NO} = 33.2$, $SD_{NO} = 15.4$, $M_{O} = 27.0$, $SD_{O} = 12.4$, $p = .009$). Analysis of mean values in the TLX subscales yielded a significantly lower temporal demand and a lower value in the performance subscale when objects were provided (temporal: $M_{NO} = 30.5$, $SD_{NO} = 21.2$, $M_{O} = 20.31$, $SD_{O} = 13.6$, $p = .009$; performance: $M_{NO} = 38.0$, $SD_{NO} = 21.4$, $M_{O} = 30.9$, $SD_{O} = 21.9$, $p = .049$).

Object positioning task

For *communication behavior*, the Wilcoxon signed rank test revealed that deictic speech was used significantly less frequently when objects were provided ($M_{NO} = 19.7$, $SD_{NO} = 4.8$, $M_{O} = 14.9$, $SD_{O} = 6.3$, $p = .036$) (see Figure 4).

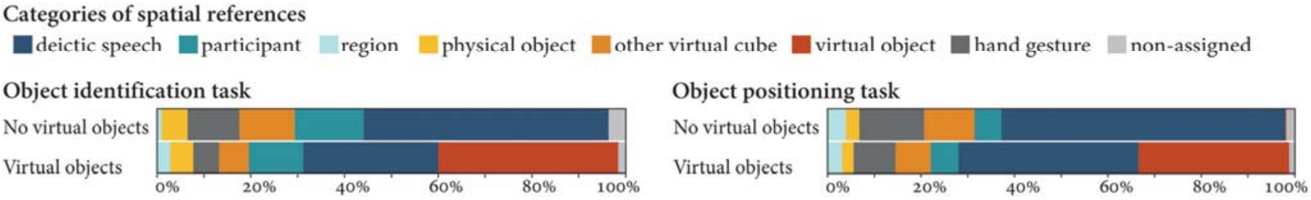

Figure 4. Categories of spatial references (top) and the mean proportions of occurrences in each study task and each condition.

For *user task load*, a pairwise comparison in the NASA TLX neither revealed a significant difference in the overall user task load nor in any subscale.

Recall precision was defined as the minimum error of the distance of positioned cube pairs compared to their former position in the object identification task. A Friedman test revealed that the minimum error was significantly lower when additive virtual objects were provided ($\chi^2(2) = 4.9$, $\bar{x}_{NO} = 1.59$, $\bar{x}_{O} = 1.41$, $M_{NO} = 3.02$, $SD_{NO} = 1.32$, $M_{O} = 2.62$, $SD_{O} = 1.17$, $p = .032$).

Further results from the concluding interview

When asked how well target objects could be identified, from 0 (very poorly) to 10 (very well), the condition with additive objects reached a significantly higher rating ($t(31) = 9.031$, $M_{NO} = 5.34$, $SD_{NO} = 1.84$, $M_{O} = 7.91$, $SD_{O} = 1.15$, $p = .000$). 10 participants reported to have perceived the physical environment to a very limited degree and other 10 reported to have not consciously perceived it at all. When additive objects were provided, they were accepted as part of the 3D interaction space, e.g., *"the room looks strange, now that all these objects are missing,"* or *"this cube was where the snack machine stood earlier."* (No additive objects were provided at the time the statement was made).

DISCUSSION, IMPLICATIONS, AND LIMITATIONS

Results from communication behavior clearly indicate the positive effect of the additive virtual objects: they were used extensively as shared spatial cues to identify target cubes and the use of the less specific deictic expressions decreased significantly along with the hand gestures which often occurred together. In line with the lower task load in the object identification task and participants' statements, the additive objects supported the participants in expressing spatial references. The object positioning task yielded a similar communication behavior, but no significant differences in the task load. This can be explained by participants' statements that positioning without the additive objects often was a mere matter of guessing, what indicates that participants put only little mental effort in recalling their positions. This aspect, however, shows that results in workload must not be explicitly attributed to communication behavior but also include individual spatial memory which might have been profited from the presence of additive objects.

Nearly all participants reported that they did not regard the physical environment in general. By some, this was explained as being because the data objects were obviously artificial and that the additive objects appeared similar in contrast to the physical environment. This clearly shows how important the additional objects were for navigation and communication in the MRE. Yet, this also raises the following question: Would the physical environment have gained more attention if the virtual objects (both the cubes and the additive objects) would have looked and behaved more like physical objects (e.g., realistic rendering, no hovering in space, etc.)? In addition, as perception within an MREs presumably depends on the proportions between the physical environments and the virtual content, communication behavior may depend on the proportions we used and the physical features of our lab. Similarly, this applies to our choice of virtual objects: other objects might have produced other results. Furthermore, and as stated by Kiyokawa et al. [8], the type of display technology is crucial in MR. The tablets required participants to focus on a rather narrow field of view. Thus, collaborators' attention was set on a small portion of the environment. Our findings may therefore not be fully generalizable as they might be specific to task design and the design of our MRE. Possible influencing factors (e.g. proportion of virtual and physical content, type and degree of abstraction of present objects, and display technology) and are subject of further research.

In the analysis of communication behavior we identified several categories of spatial expression that participants made use of. These categories only partially overlap with those identified by Chastine et al. [4]. Hence, we do not claim the classification as complete but rather task specific. Furthermore, our classification did not take the complexity of natural language into account. This should be further investigated from a linguistic point of view. Logan [10] and Levinson [9], for example, provide a deeper understanding of cognition and space in language that focuses on the rather complex relationship between 1) the observer, 2) the addressee, 3) the target object (here: a target cube), 4) the referenced object, and 5) several types of frames of reference. One interesting aspect refers to the concept that objects provide their own intrinsic coordinate systems that differ in their expressiveness (e.g., a chair has its clear front side, a bottle without a label does not). Thus, expressiveness of an object's intrinsic coordinate system may further help to disambiguate spatial expressions in MREs and should therefore be studied.

CONCLUSION

This note contributes to a better understanding of the effects of spatial cues in collaborative MREs. Results from our study, consisting of an object identification task and an object positioning task, provide three major insights: 1) The physical environment plays only a minor role in the collaborators' perception and their communication behavior, whereas 2) the virtual objects are extensively used as spatial cues by collaborators, and 3) their presence positively influences collaborators' communication behavior, decreases user task load, and improves user experience. This suggest adding virtual objects to MREs to reduce user task load and improve groups' communication behavior and user experience.

ACKNOWLEDGEMENTS

We thank the German Research Foundation (DFG) for financial support within project C01 of SFB/Transregio 161. We also thank VIS Games, Jason Wong, and Vertex Studio for providing us the Unity assets we used as additive virtual objects.

REFERENCES

1. Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. 1998. Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries. *ACM Transactions on Computer-Human Interaction* 5, 3: 185–223. http://doi.org/10.1145/292834.292836
2. Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2: 73–272. http://doi.org/10.1561/1100000049
3. Mark Billinghurst and Hirokazu Kato. 1999. Collaborative Mixed Reality. *Proc. of ISMR '99*, Springer, 261–284.
4. Jeffrey W. Chastine, Kristine Nagel, Ying Zhu, and Luca Yearsovich. 2007. Understanding the design space of referencing in collaborative augmented reality environments. *Proc.of CGI '07*, 207–214. http://doi.org/10.1145/1268517.1268552
5. Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of Communication : Effects of Shared Visual Context on Collaborative Work. *Proc. of CSCW '00*, 21–30. http://doi.org/10.1145/358916.358947
6. William W. Gaver. 1992. The Affordances of Media Spaces for Collaboration. *Proc. of CSCW '92*, 17–24. http://doi.org/10.1145/143457.371596
7. Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2013. Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Human-Computer Interaction* 28, August 2014: 1–39. http://doi.org/10.1080/07370024.2012.678246
8. Kiyoshi Kiyokawa, Mark Billinghurst, Sohan Hayes, Arnab Gupta, Yuki Sannohe, and Hirokazu Kato. 2002. Communication Behaviors of Co-Located Users in Collaborative AR Interfaces. *Proc. of ISMAR '02*: 135–144. http://doi.org/10.1109/ISMAR.2002.1115083
9. Stpehen C. Levinson. 2003. *Space in language and cognition: Exploration in cognitive diversity*. Cambridge University Press. http://doi.org/10.1017/CBO9780511613609
10. Gordon D. Logan. 1995. Linguistic and Conceptual Control of Visual Spatial Attention. *Cognitive Psychology 28*: 103–174. http://doi.org/10.1006/cogp.1995.1004
11. Paul Milgram and Fumio Kishino. 1994. Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 1321–1329. http://doi.org/10.1.1.102.4646
12. Daniel R. Montello. 2005. Navigation. In *The Cambridge handbook of visuospatial thinking*, Pritt Shah and Akira Miyake (eds.). 257–294. http://doi.org/10.2277/0521001730
13. Kjeld Schmidt and Carla Simone. 1996. Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)* 5, 2-3: 155–200. http://doi.org/10.1007/BF00133655
14. Norman G. Vinson. 1999. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. *Proc. of CHI '99*, 278–285. http://doi.org/10.1145/302979.303062
15. 2015. Project Tango Tablet Development Kit User Guide. Retrieved January 3, 2016 from https://developers.google.com/project-tango/hardware/tablet#technical_specifications
16. 2015. Unity – Gaming Engine. Retrieved January 3, 2016 from https://unity3d.com/