Collaborative Identification of Objects in Physically Separate Mixed Reality Environments

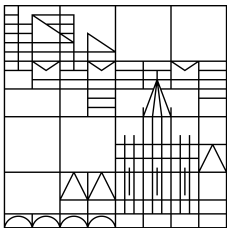
Bachelorarbeit

vorgelegt von

Matthias Kraus

an der

Universität Konstanz



Human-Computer Interaction

Fachbereich Informatik und Informationswissenschaft

1.Gutachter: Prof. Dr. Reiterer

2.Gutachter: Jun.-Prof. Dr. Gipp

Konstanz, 2016

Abstract

Prior research provides initial evidence that virtual objects can serve as spatial cues and can hereby have a favorable effect on the collaboration behavior of co-located participants in mixed reality environments (MREs), i.e., newly created environments composed of coexisting real and virtual objects. To date, the influence of virtual objects on collaboration in distributed MREs - where people are situated in physically separate locations while at the same time sharing one and the same virtual blending - has been a point of minor scholarly interest. To address this research gap, we investigated how additional virtual objects shape collaboration in object identification tasks in remote MREs. For this purpose, a controlled lab experiment with 16 dyads was conducted. Results indicate that spatial cues can affect user experience beneficially. This is apparent in reports of participants, which uniformly preferred the condition with virtual spatial cues. Further advantages of these cues can be seen in the facilitation of communication as well as in significantly affected experiences of social presence, both being gathered on the basis of questionnaires. These findings emphasize the usefulness of synthetically created spatial cues for collaborative object identification tasks in remote MREs. Possible application areas constitute future office environments, where this kind of spatial cues could help co-workers to mutually undertake a project while being physically separated.

Table of Contents

List of Figures

List of Tables

Abbreviations

ADF Area Description File

AR Augmented Reality

CSCW Computer Supported Cooperative Work

CSV Comma-Separated Values

GUI Graphical User Interface

MR Mixed Reality

MRE Mixed Reality Environment

RPC Remote Procedure Call

SDK Software Development Kit

TLX Task Load Index (measures workload)

TPI Temple Presence Inventory (measures cognitive presence)

VR Virtual Reality

XML Extensible Markup Language

1 Introduction

The blend of two separate, de facto independent, environments from which one is real and the other synthetic can be referred to as a Mixed Reality (MR; Milgram and Kishino, 1994). In this context, synthetic means virtually created. Figure 1.1 demonstrates an example for blending a virtual environment (plants and a bookshelf) into a real environment, thereby creating an MRE. It is important that each virtual object is registered to a certain physical location (i.e., the virtual objects keep their position continuously). Thus, a consistent composition of real and virtual objects is created. In case the virtual component is rendered photo-realistically, the observer of this environment might even be incapable of distinguishing between reality and surreal add-ons. MREs can be observed using different mediums like monitor based video displays, handheld displays or head-mounted displays. Partly dependent on the chosen AR display, a suitable AR interface can be used for user input. During the past decades, five main types of AR interfaces were developed: *Information Browsers*, *3D User Interfaces*, *Tangible User Interfaces*, *Natural User Interfaces* and *Multimodal Interfaces* (Billinghurst et al., 2015).

Figure 1.1: A MR: Virtual objects are blended into the real environment.

In order to classify and to distinguish between different types of MRs, Milgram and Kishino (1994) introduced the virtuality continuum (see Figure 1.2). It describes the transition from real to virtual environments or vice versa passing different forms of MRs. The term mixed reality is a hypernym of all kinds of composites of real and virtual environments. One can distinguish between different types of blendings, for instance, on the basis of the original frame of reference. E.g., while augmented reality environments are based on the real environment, augmented virtualities extend virtual environments by real elements (Figure 1.2, AR/AV). In MREs a person's visual context can be separated into two different parts: the virtual and the real component.

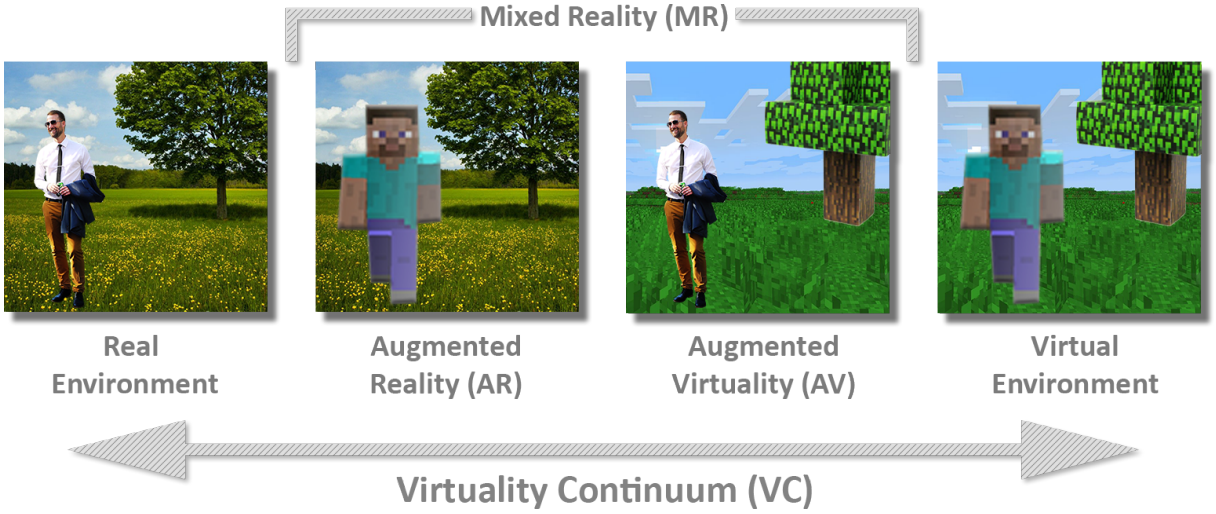


Figure 1.2: Virtuality Continuum (Milgram and Kishino, 1994).

In the following, the term semi-shared reality is used for scenarios in which only a few parts of the collaborators' environments are shared. An intuitive example for this is a remote collaboration scenario, in which collaborators share their virtual component completely (e.g., virtual flip board), while at the same time each collaborator has his/her own real component (physical environment). In this context *remote* means that collaborators are located at different physical locations. Figure 1.3 shows the basic concept of semi-shared MRs by use of a shared virtual flip chart.

One conceivable application domain of MREs is the workplace. In office environments, digital information and tools such as pin boards (demonstrated by Kato et al. (1999)), conceptual 3D models (Golparvar-Fard et al., 2009) or even holograms of remote co-workers (Prince et al., 2002) could be inserted. MREs would offer people, who are located at different places, the opportunity to join the same semi-virtual environment, hereby providing a shared workspace. This way, coordination and communication efforts could be eased substantially.

To illustrate this point, one can think of two architects living in distinct cities but working on the same project, for example, the design of a building. Certainly, the collaborators

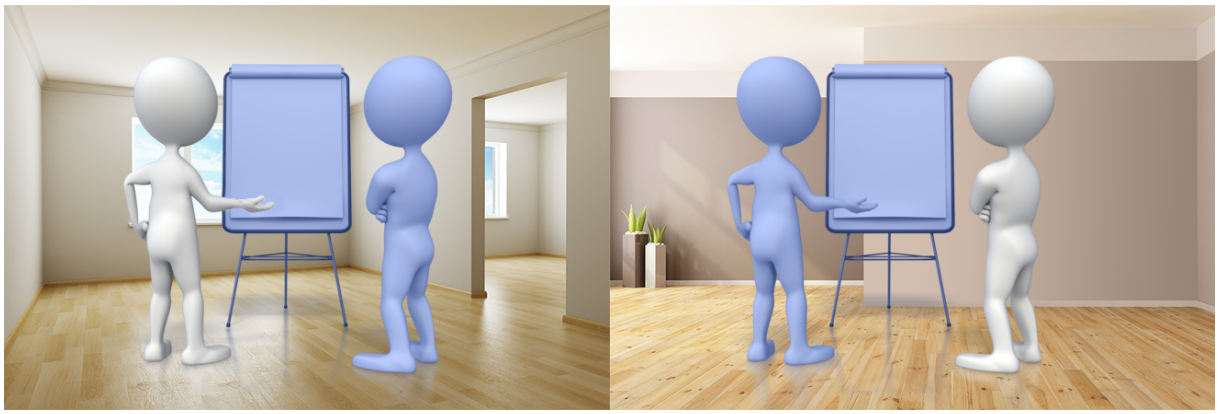


Figure 1.3: A semi-shared MR. The virtual component (blue) consists of a virtual flip chart and a model of a remote collaborator. Each collaborator has his own physical environment.

could meet face-to-face whenever it is necessary to proceed with the work. But, with the aid of computer technology, remote collaboration would also be possible. For instance, an architectural 3D model of their work could be created in order to then be displayed in their respective environments. The two architects could work simultaneously on the project, thereby allowing everybody to stay where he or she is. Instead of only manipulating the 3D model on the architects' computers, the model could be displayed in an MRE. This would not only enable each person to interact with the model, but also to recognize what the other is doing. One of the major obstacles for efficient remote collaboration is the dissimilarity of collaborators' physical surroundings (Clark and Brennan, 1991). The aspect of having a mutual frame of reference, of which both involved parties are able to make use of, is crucial (Clark and Brennan, 1991). Therefore, creating a common virtual environment would help to overcome the physical remoteness between the two architects or other collaborating individuals.

Another useful application domain of MREs is the sharing of one's real environment with others for collaboration purposes (principle depicted in Figure 1.4). Instead of creating a new virtual environment which both collaborators enter (Fig. 1.3), the real surrounding of one collaborator is projected to the other one. To facilitate natural face-to-face communication, a 3D model of the remote co-worker is embedded in the respective local environment. Additionally, virtual elements can be inserted if necessary.
In this case, the common frame of reference is no longer an additional layer (newly created virtual world), but the real environment of one collaborator which is transmitted to the other collaborator. A person's real surrounding, or at least parts of it, could be virtually embedded in the environment of the other person. Conversely, the second person is virtually added to the first person's environment. Thus, creating the impression that both are located in one room. A vivid example of this kind of application is a scenario with a futuristic plumber. The previously presented technique could facilitate the plumber to enter the client's bathroom

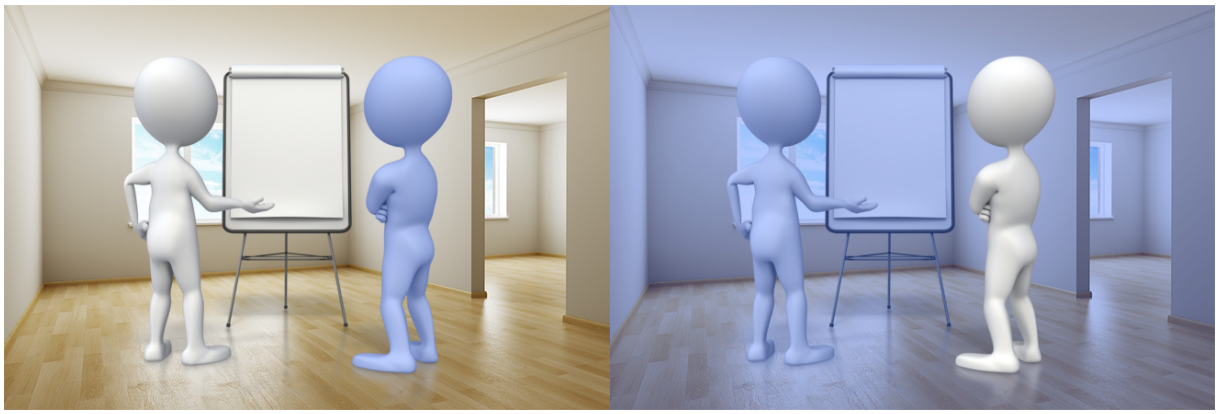


Figure 1.4: The virtual components are displayed in blue. The physical environment (room, flip chart) of the first collaborator (left) is completely shared with the second co-worker (right). The first collaborator interacts in a completely shared MR, whereas the second person interacts in a completely shared VR.

in order to inspect a malfunctioning sink. From the client's point of view, a hologram-like representation of the plumber would appear in the bathroom, whereas the plumber, on the other side, sees the client and a part of his bathroom including the sink. He could properly inspect it and tell the client how to fix it by pointing on parts of the pipes or even by marking them with an additional spatial marker.

Microsoft presented a similar scenario in one of their HoloLens[1] promotion videos. Figure 1.5 shows two snippets from the footage. In their scenario a father explains his daughter how to fix a sink with the aid of visually embedded 3D markers in the daughter's MR. The main difference to the aforementioned example is that their approach restrains the father's view to the daughter's field of view. In this scenario only the daughter has a "real" interactive MRE, whereas the father only sees a static video stream (i.e., he cannot take a look around as his view is bound to the perspective of the camera in his daughter's device).

The same technique could also be used for instructors or teachers. E.g., a fitness instructor could join a private gym and explain the proper usage of individual gym equipment by pointing to things directly. Moreover, he could walk around and observe the whole training process from different angles. The list of potential applications for this technique could be continued endlessly, as there are plenty of promising ambits.

The mentioned notions imply that it is essential that all users in MREs indeed relate to the surrounding they share - no matter if virtual or real. Otherwise, communicational difficulties might arise as each person is exposed to a different setting. For instance, if one architect from the first example uses the real environment to tell his co-worker where he should change something on the model, it can quickly be very confusing as the same point of reference

[1]Microsoft HoloLens (2016) - head-mounted MR display.

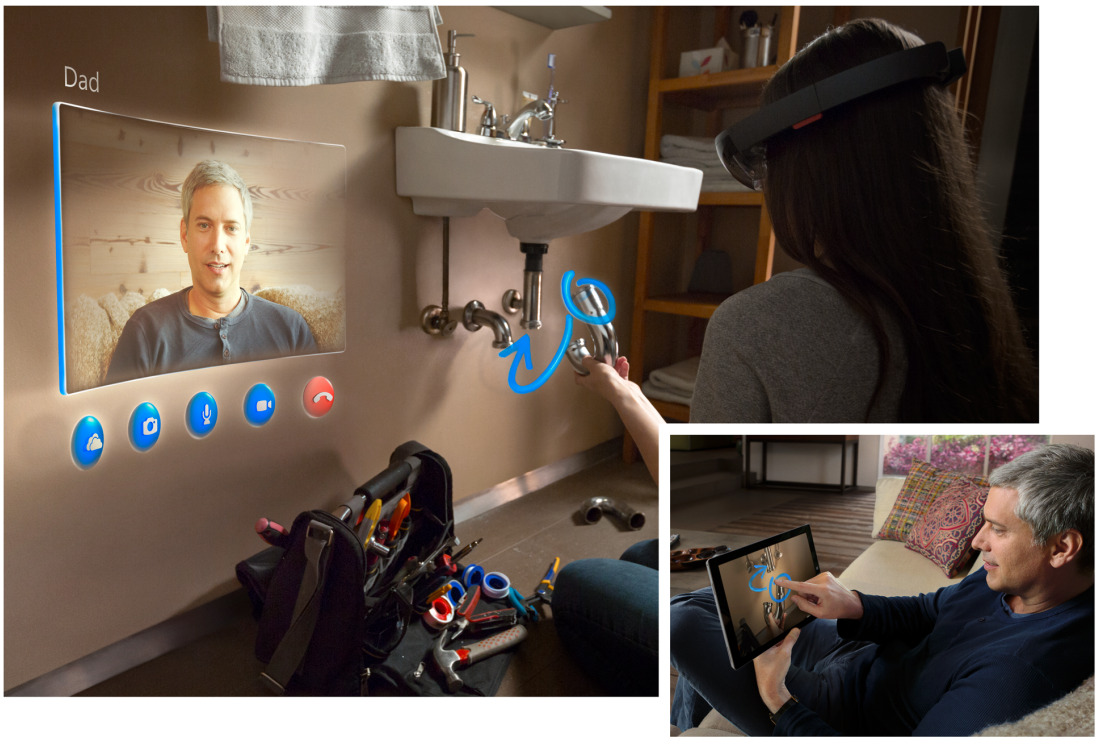


Figure 1.5: Microsoft HoloLens (2016). Promotion video footage. The MR scenario: A father explains his daughter how to fix the sink with the aid of visual markers integrated into the daughter's real environment.

(e.g., door, window, desk, ...) could be somewhere else or does possibly not even exist in the second architect's environment. Thus, completely shared environments (i.e., MR and VR; Fig. 1.4) are - in terms of conversational grounding - advantageous compared to semi-shared MREs (Fig. 1.3). But, with regard to orientation, navigation, and user experience, MREs provide a better individual basis to work in than virtual reality environments (Billinghurst and Kato, 1999). This is mainly because see-through displays offer more natural navigation and interaction possibilities with the physical environment. Therefore, a combined solution is required, which provides both - a great common conversational grounding similar to the one in totally immersive virtual environments and the advantages of MREs (orientation, navigation, natural interactions with the physical environment and its tools).

As shown, MREs can facilitate cooperation between individuals. The application of MREs as a supportive tool for remote collaboration and other remote cooperative activities becomes more and more suitable for private consumers and industrial end users. According to Clark and Brennan (1991), one of the biggest problems related to efficient and natural collaboration is the lack of a common visual grounding in remote MREs. For this purpose,

this thesis examines whether and how an additional shared virtual context influences the collaboration of people located in remote physical environments. Specifically, the paper focuses on the impact of digitally created visual cues on user task load, communication behavior, subjectively perceived presence, and user experiences on object identification tasks in remote co-located MREs.

In the following, several topics of related work will be reviewed and discussed in order to introduce basic principles and ideas for the subsequent sections. Furthermore, the next passages serve to outline our motivation behind the conducted experiment by presenting use cases and previous studies with similar application domains and research questions. After that, our research prototype for the study is presented, including its basic concepts and its development process. This is followed by the main part of the thesis - the study itself - including study design, conduction, results, and evaluation. As a bottom line, a conclusion wraps up the overall achievements of the conducted study.

2 Related Work

This chapter provides an overview of the most important related strains of research. To begin with, different basic applications of mixed reality environments will be presented. Afterwards, the collaborative aspect of MREs is taken into account. Then, the focus is put on *remote* computer supported cooperative work (CSCW). Last but not least, several publications, which dispute the employment of virtual spatial cues in CSCW-environments, are briefly reviewed.

2.1 Working in Mixed Reality Environments (MREs)

Milgram and Kishino (1994) define "Mixed Reality" as the simultaneous display of real objects in combination with virtual components on one single device. Thus, a MR is the blend of real and virtual environments. Furthermore, they describe several factors by which different types of mixed reality displays can be distinguished. The authors identified three major influencing factors: *Extent of World Knowledge*, *Reproduction Fidelity*, and *Extent of Presence Metaphor* (Figure 2.1). The dimension *Extent of World Knowledge* indicates how much is known about the virtual world. It describes which and how many objects exist, where (relative to the real environment) they are located, and if the information they provide is complete (e.g., whether they can be inspected from all angles / 3D-information exists). Hereby, it can often be determined if the respective application can be numbered along the class of augmented reality applications (parts of the virtual world are known and displayed) or whether it falls within the domain of virtual reality applications (complete world is known and displayed; i.e., total overlay). Apart from this, there are applications which cannot be categorized in either of the two classes - for instance, the display of a single object without a fixed position in the real environment. In that case, the virtual blend neither represents a virtual reality, nor does it extend the real environment by adapting to it seamlessly (augmented reality). By definition of MRs, virtual objects within the MR have to be registered to a certain position in the physical environment. Therefore, these applications could be classified as 'no MRs'. The second factor described by Milgram and Kishino (1994), *Reproduction Fidelity*, takes into consideration to which extent the displayed virtual objects are considered as realistic. It measures "the quality with which the [synthesizing] display is able to reproduce the actual or intended images of the objects being displayed" (Milgram and Kishino, 1994). The third dimension *Extent of Presence Metaphor* provides information about how "present" one feels while being in the MR. This depends on several factors, such as what kind of display is used or how realistic the virtual objects are.

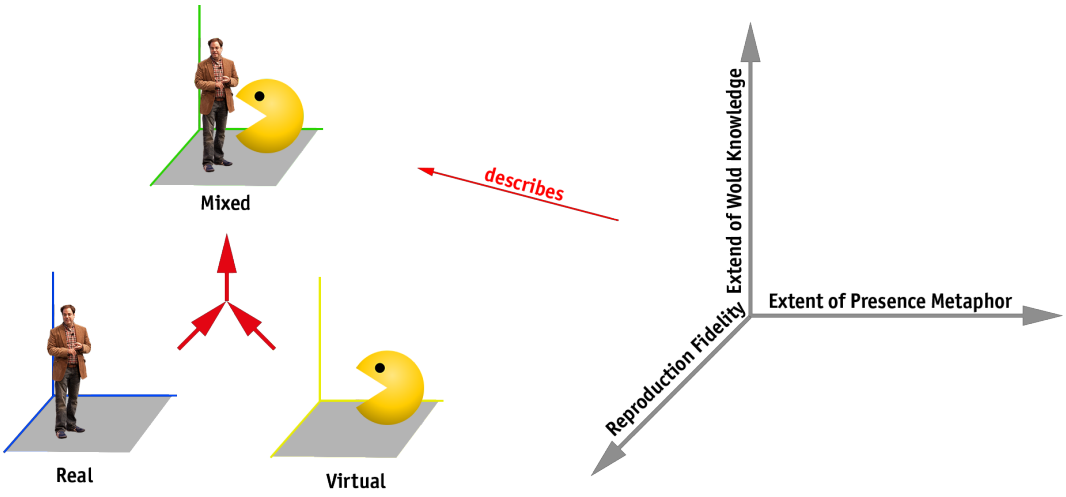


Figure 2.1: The blend of real and virtual environments (left) is called Mixed Reality. Its characteristics can be described by three dimensions (right). (Milgram and Kishino, 1994)

The idea to embed virtual elements in the real environment opens up new possibilities to enhance work procedures and personal activities. For instance, it would be possible to implement edge detection for head-mounted optical displays, such as "Google Glass" (Google Glass Developer), and to overlay the edges on its see-through display (Hwang and Peli, 2014). This technique could be further developed in order to enhance everyday life by highlighting only relevant edges like the first step of a stairway, the curbside or low hanging branches on a bike lane.

Another example out of the large repertoire of use cases was proposed by Golparvar-Fard, Peña-Mora and Savarese (2009). They introduced a framework for displaying 3D construction plans in construction sites (see Figure 2.2). A virtual 3D model of the finished work process (e.g., main girder basement) is, due to high processing effort, only integrated into 2D images of the construction site. Different parts of the model are then colored with respect to their completion status. For example, if one part is behind schedule, it is colored in red, whereas the remaining parts appear in green. This model might be further developed in order to be applied in real time and to be used to display the current work progress on a head-mounted display worn by the chief of construction.

MRs could be utilized in learning environments as well. As shown by Pan et al. (2006), the usage of MRs has great potential in the context of learning activities. With the help of MRs, topics could be conveyed in an interesting way to learners. Moreover, immediate feedback could be employed to instantly reflect what has already been learned and to monitor the learning progress. This could lead to a better "active" learning process. Pan et al. (2006) offer several promising studies and use cases in this field of application.

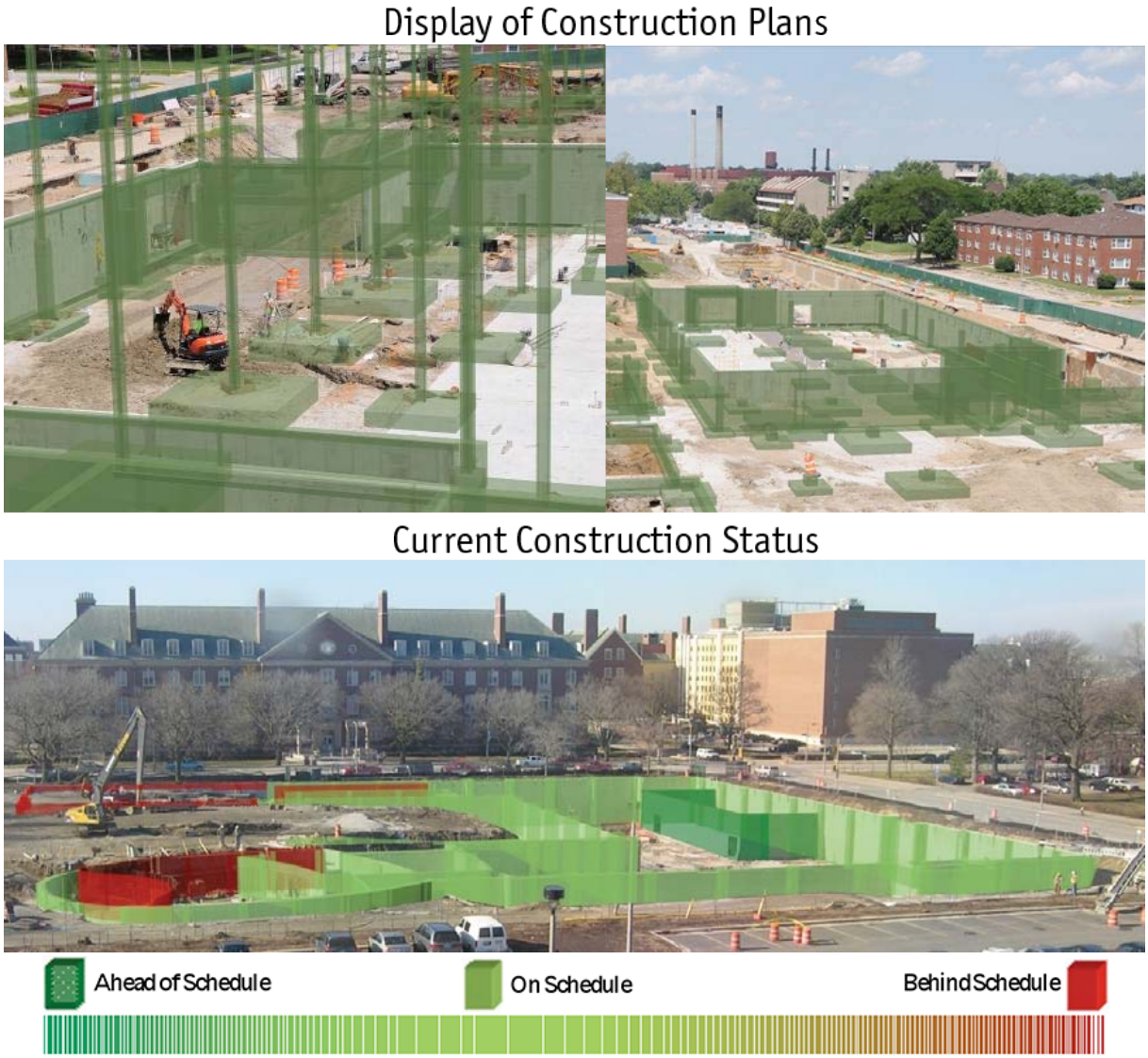


Figure 2.2: Prototype by Golparvar-Fard et al. (2009). MRE for the visual display of construction plans in a construction site (top) and visual highlighting of the current progress of the construction (bottom).

2.2 Computer Supported Cooperative Work (CSCW)

In their paper, Kiyokawa et al. (2002) dealt with the communication behavior in co-located collaborative AR environments and how it changes with various types of displays. Dyads were instructed to complete an object identification task using four different displays. During the task, participants sat on opposite sides of a desk and cooperatively identified specific virtual objects among other similar objects, which were virtually embedded upon the table. The four distinct types of displays were: optical see-through head-mounted display (HMD), video see-through HMD, monoscopic see-through (2D video stream and 3D cube information), and virtual reality display (showing only the cubes and the opponent's face in a virtual environment). They came to the conclusion that "optical see-through [...] was the best in terms of words and gestures needed to complete the task" (Kiyokawa et al., 2002). Moreover, they hold the view that - contrary to their results - the differences between optical and virtual see-through displays are negligible. They explain that the lower scores using the virtual displays were caused by the camera offsets in their setting.

Several researchers have focused on the question whether collaboration can somehow be improved by using mixed realities instead of pure virtual environments in collaboration tasks. According to Billinghurst and Kato (1999), the use of mixed realities provides a number of benefits. Face-to-face MREs would not only make it easier for users to retain their communication and interaction habits using glances, gestures or facial expressions, but also would it allow users to keep their physical awareness. Hence, they would be able to integrate their real surroundings into their perception.
Another team facing the same question conducted a case study using a game as a basis. Groups of two were instructed to play air hockey - once in a virtual reality and once in an augmented reality environment (Ohshima et al., 1998). The evaluation of their study revealed that their "experimental collaborative AR system achieve[d] higher interactivity than a totally immersive collaborative VR system" (Ohshima et al., 1998).

Another interesting paper by Billinghurst and Kato (2002) addresses a similar topic, namely collaboration in augmented realities and its advantages over VR or screen-based methods. They compared multiple, previously conducted studies. Among other things, they came to the conclusion that AR-techniques have the *unique* ability to enhance the reality by facilitating seamless interactions between real and virtual environments. That is, only AR-methods would be able to get close to face-to-face like collaboration scenarios.

A good example for a MR-based CSCW-environment is an innovative learning space for elementary school children. They could interact within this environment, while using tools and information provided by the MR at the same time. A prototype for this kind of application was developed by Kritzenberger et al. (2002). In their study, they investigated pupils' interactions as they built their own MR-landscapes using virtual and real elements.
Another imaginable use case was presented by Prince et al. (2002). In their paper, they

put forward a novel system, which facilitates hologram-like video-conferencing. Using a head-mounted see-through MR-display, the person called can be virtually inserted into one's real environment. This allows people to speak authentically face-to-face, despite being far away from each other. Using multiple video cameras, the person on one end is scanned and completely modeled in 3D. This model is then rendered into the head-mounted MR-display of the person on the other side. This technique could not only be used for video conferencing, but also for collaborative, educational or entertainment purposes. Similar approaches were accomplished by Billinghurst at al. (2000, 2002).
One last example, provided by Benko et al. (2004), is the setup of a semi-virtual archaeological excavation site. A 3D model of such a site can be displayed within an office environment, supplying a foundation for discussion and exchange for multiple researchers (Figure 2.3).

Figure 2.3: Archaeological excavation site (left) displayed in MR within an office environment (right). (Benko et al., 2004)

2.3 CSCW in Physically Separate Environments

According to the taxonomy introduced by Benford et al. (Benford et al., 1998), a "shared space" can best be described by three dimensions - *transportation*, *artificiality* and *spatiality* (see Figure 2.4). The first dimension - *transportation* - comprises "the degree to which users are transported into some new space or remain in their local space" (Benford et al., 1998). *Artificiality* describes to which degree the shared space is based on the real environment and how much of it is created synthetically. The last dimension, *spatiality*, "concerns the degree to which the shared space exhibits key spatial properties such as containment, topology, movement, and a shared frame of reference" (Benford et al., 1998). Figure 2.4 illustrates these three dimensions employing several examples of different shared spaces.

With this taxonomy, the semi-shared MRE, which we consider in this thesis, can best be located in the diagram of Figure 2.4 between blue and yellow. Two collaborators join the

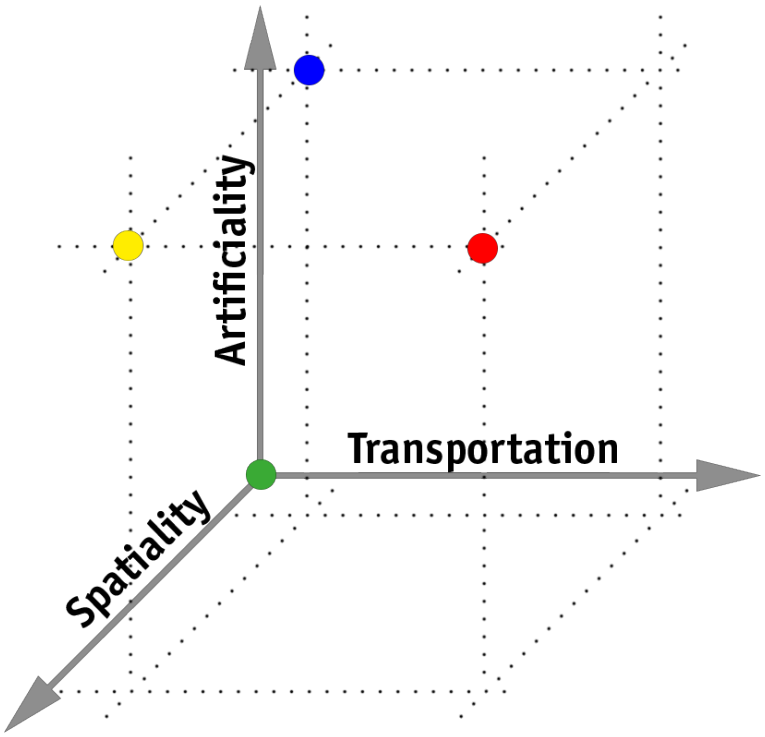


Figure 2.4: Taxonomy of shared space by Benford et al. (1998). Dimensions which define the characteristics of a shared space. Green: physical reality, no shared frame of reference (spatiality), no transportation, no synthetic objects (artificiality). Blue: augmented reality with synthetic objects (artificiality), no longer completely locally (transportation), small frame of reference - e.g., collaborators share only few virtual elements (spatiality). Yellow: augmented reality (see blue) with common frame of reference (collaborators in same room + shared virtual contents). Red: shared virtual reality, complete transportation and artificiality, complete VR is shared (spatiality).

same virtual component of the MR, whereas the physical part is individual (see Introduction Fig. 1.3). Additionally added virtual cues change the degree of artificiality and spatiality in this model, as they are shared virtual components. Low values in the dimension of transportation are due to the fact that most of the persons' environment is real and that they are still aware of their physical surroundings.

As reported by Benford et al. (1998), mixed realities are shared spaces as soon as more than one person is present in the same room, because MREs combine a physical space (which is shared unavoidably) and a virtual environment. In their paper, they present a model with so-called "mixed-reality boundaries". Instead of overlaying virtual objects on top of the real environment - like in the model by Milgram and Kishino (1994) - they intended to build window-like boundaries between the two worlds, basically linking the two worlds completely

to each other. To illustrate their model, they implemented a mixed reality boundary by joining a website to a real place (a foyer). Internet users were able to join the shared space seen by other virtual users (on the website) and by the physical visitors (on a wall mounted screen) as 3D icons. This application enabled them to communicate with each other and, at the same time, with real guests. A similar prototype was introduced by Brown et al. (2003). In their example, web and physical users were able to visit a museum at the same time. Thereby, they were aware of each other and could communicate and even interact. Every physical visitor had a portable device on which he/she could see virtual visitors.

Fussell et al. (2004) investigated in more detail the role of gestures in remote collaborative tasks. They compared the effects of real gestures in co-located scenarios with cursor-based pointing devices in remote settings. Furthermore, they introduced a pen-based pointing device which was capable of conveying the meaning of the remote gesture to a greater extent. They found that cursor-based pointers were not sufficient for remote collaboration. Therefore, other technical methods are needed to convey visual information, such as gestures, in order to receive a scenario in which work can be performed in a similarly efficient way as in co-located environments.

Several previously named examples of mixed reality or collaborative applications could also be employed on remote collaboration tasks. For instance, video conferencing, extended by some features, offers the possibility to interact. Kato et al. (1999) developed a 3D conferencing application, which additionally included a shared virtual white-board. All participants of the conference could be placed as "windows with video streams" arbitrarily in the room, still remaining a 2D canvas each. Each participant was able to draw onto another virtual 2D canvas, which was shared amongst all of them. A second example for computer supported remote collaboration is a setup introduced by Robinson and Tuddenham (2007). Teams of multiple users collaborated at touch-sensitive tables in physically remote locations. In this mixed *presence* environment, remote contributors were displayed as shadows on the screen, meaning everybody saw the hands of all remote collaborators at their respective position as 2D representations on the touch table.

2.4 Virtual Spatial Cues in CSCW-Environments

Another key aspect of this thesis is the examination of how shared visual context influences collaborative work. Fussell et al. (2000) investigated the impact of a shared visual space on collaborative performance. In their study, dyads had to solve a collaborative manual task. Three different scenarios were tested. In the first task, both participants were co-located to solve a collaborative repair task. In the second one, they were physically separated and were only able to speak to each other. The third scenario offered video transmission in addition to auditory communication. Even though the collaboration performance with video and audio combined exceeded the performance attained given solely auditory communication

possibilities, it could not match up with the co-located collaboration. As reasons for that, they name conversational grounding and the point that visual information could facilitate the instantiation of grounding in many different ways. Furthermore, the physical constraints of the video camera and its setting is a possible restraint for the common grounding (e.g., if a relevant part of the shared space is not transmitted).
The principle of common conversational grounding is explained by Clark and Brennan (1991). First, both dialog partners have to meet the *grounding criterion* (Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986). The criterion is chosen in regard to the respective situation and describes the minimum degree to which a contributor has to think that his partner understands sufficiently what he is trying to convey of his respective situation. "[Conversational] grounding is [then] the collective process by which the participants try to reach this mutual belief" (Clark and Brennan, 1991). In other words, one tells the other something he would probably understand, and then tries to verify whether or not he understood. Once a level is reached on which both can communicate, a grounding is set and no further verification is necessary. For instance, if a person intends to tell another person that he sees a nice car and the other person is standing right next to him, it is sufficient to say "nice car!" (the second person would see the car as well and realize what he meant). Not so if the other person is on the phone. In that case, the one talking has to figure out what he needs to say in order to make the other person understand what he means (e.g., "here is a nice car!"). Once the situation is clarified, he could also use expressions like "decent carbon rims!" - and the other one will still realize that he means the car's rims. A common grounding is set.

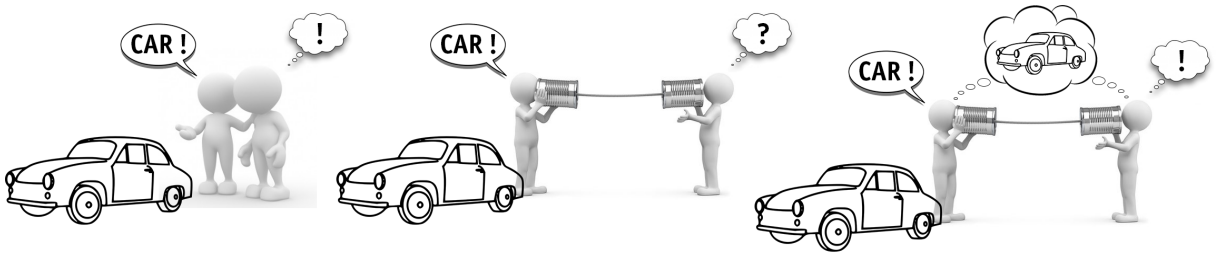


Figure 2.5: Common conversational grounding is crucial for effective communication. (Physical) Environments can provide such a grounding (left). In case not all participants of the communication are on the same level (center - remote locations, different environments), they first have to establish a common grounding (right). This does not only hold for environmental references, but also for different (non-material) topics like scientific discussions (i.e., definitions and notations have to be exchanged in advance).

Multiple examples in the previous section have already indicated that a common conversational grounding can be crucial for good collaborative results. Gergle et al. (2013) tried to examine this topic more precisely with the help of a series of experiments. Their overall conclusion reads as follows: "Visual information about a partner and the shared objects that

comprise a collaborative activity provides many critical cues for successful collaboration" (Gergle et al., 2013). According to them, situation awareness in combination with conversational grounding is the key for good results in remote collaborative tasks.

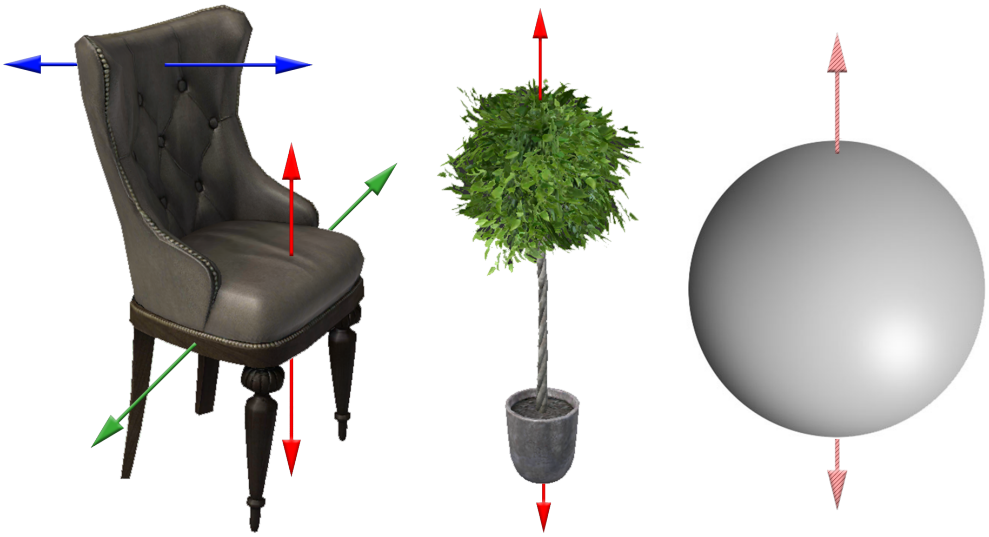


Figure 2.6: Orientation of visual objects. Some objects only provide one absolute (independent of observers' position) referable dimension (center, red: "on top" or "below"), while others have a more specified orientation (left: 3 dimensions - red ("on top"), green ("right of"), blue ("behind")). Other objects, like a sphere (right), do not have an own orientation and can at most be absolutely referenced by their static position in the room (top, bottom does usually not change with perspective). Vinson (1999).

As several MR applications merely have a small amount of shared objects, it might be reasonable to add - although for the respective task itself dispensable - virtual objects in order to strengthen users' common conversational grounding (more objects to reference) and situation awareness (eases imagination for virtual component of the MR). Norman Vinson (1999) deployed a basic set of guidelines on how to support navigation in virtual environments by additional visual landmarks. He justifies his findings with examples and similarities in "extensive empirical literature on navigation in the real world" (Vinson, 1999). In his opinion, landmarks, which provide an orientation (front, side) instead of solely a position (e.g., sphere), serve as especially suitable spatial cues (see Figure 2.6). The same applies to visual cues that are particularly noticeable and outstanding. Even though Vinson only refers to single user navigation, his guidelines can be transferred to collaborative work environments - even to those, which are not completely virtual - as the guidelines should solidify the spatial awareness among users in these settings as well. Again, the higher level of spatial awareness would probably lead to better collaboration results (Gergle et al., 2013).

The research work by Müller et al. (2015) provides a collaborative scenario with an object identification task and additional, synthetic spatial landmarks. They examined the influence of virtual cues on users' communication behavior and their task load. Thereby, they came to the conclusion that the enrichment of mixed realities with additional spatial cues positively affects users' communication behavior, enhances their user experience and lowers their overall task load. As the content of this thesis is very closely related to the one provided in Müller et al.'s paper, we will take a closer look at several parts of their publication throughout the following chapters.

The content of the previous sections shows the wide range of application possibilities for the display of additional visual contents, with the goal to enhance daily procedures or work processes. One strain of application is remote collaboration. Given the described information, it seems comprehensibly that a common conversational grounding is important for efficient collaboration. Even though totally immersive virtual environments may provide a complete conversational grounding, they lack in other features, which MREs supply (improved navigation, orientation, user experience etc.). Therefore, a solution is required which is capable of improving collaborative work in remote MREs. One possible approach is to broaden the common conversational grounding by providing more shared virtual objects. Hence, we conducted a study in order to find out *"how additional visual cues in remote collaborative MREs influence collaboration during object identification tasks"*. In the following chapter our research prototype is presented, which was developed to investigate this issue.

3 Research Prototype

The study for this bachelor thesis was conducted in cooperation with Matthias Miller, who carried out research on a similar subject in the same domain. Both of our research questions could widely be investigated with the same test scenario. Therefore, we covered the two fields of interest with only one study consisting of two individual tasks. The conducted study was a follow-up study based on the one Jens Müller et al. (2015) described in their paper.

Their findings suggest that virtual cues, which are added to a MRE, can enhance users' collaboration in co-located environments (Müller et al., 2015). Yet, by now, it is still unclear and insufficiently studied how collaborators in physically remote settings are influenced by spatial cues. Hence, the main idea guiding our investigations was to find out whether and how the situation changes in case the collaborating members are no longer located in the same physical environment, but still interacting with the same virtual environment. More specifically, we were concerned with the research question of how virtual objects, which are added to the MRE, influence collaboration during object identification and positioning tasks.

3.1 Concept

We adapted the main idea from the study by Müller et al. (2015) and used it as a basis for our research study, since our independent variables as well as most of the dependent variables match theirs. The most noticeable difference between the two experiments is the remote setting in our case. Whereas Müller et al. (2015) focused on the examination of the cues' influence in co-located MREs, we dwelled on the influence of cues in distributed MREs. In particular, we investigated the following research question:

How do additional virtual objects shape collaboration in object identification tasks in physically separate mixed reality environments?

We opted to employ a similar setup and the same tasks as Müller et al. (2015), because this opens up the possibility to make direct comparisons between the two studies. The prototype of Müller et al. (2015) provides two different tasks, which partly depended on each other. During the first task, participants play a modified 3D version of the "memory" card game. In the second task, players are encouraged to reconstruct the constellation of cubes from the previous task. The main focus of this thesis is to examine the influence of virtual

cues on collaboration in object identification tasks (first task). The second task, consisting of an object positioning exercise, is Matthias Miller's major point of interest. Nevertheless, as the second task (positioning task) is dependent on the first task, it is possible that its results are influenced by the first task as well and are therefore also taken into consideration.

In the first task, virtual cubes are distributed everywhere in the room. They can be opened by clicking on them on the touch screens of the tablets each participant receives at the beginning of the experiment. When opened, a cube displays its texture. Analogous to the card game, once two are opened with the same texture, a match is found and the two cubes are removed from the game. If they do not match, they are closed again. The players have to search on, until all matches are found and the game terminates.
The second task starts with an empty playing field (matching memory cubes disappeared). Participants are instructed to set all cubes back to their original position (using buttons on the tablets' GUI).

Our prototype is designed for two participants, as we intended to create a minimal (remote) collaboration scenario. They are located in different rooms and each of them is equipped with a tango tablet, allowing them to join the same MR - or more precisely - the MR's shared virtual component. Figure 3.1 shows snippets of our final research prototype. Both players see the same virtual elements (cubes, plants), but each has a different physical environment. Additionally, an abstract model of the teammate's tablet is depicted (Fig. 3.1).

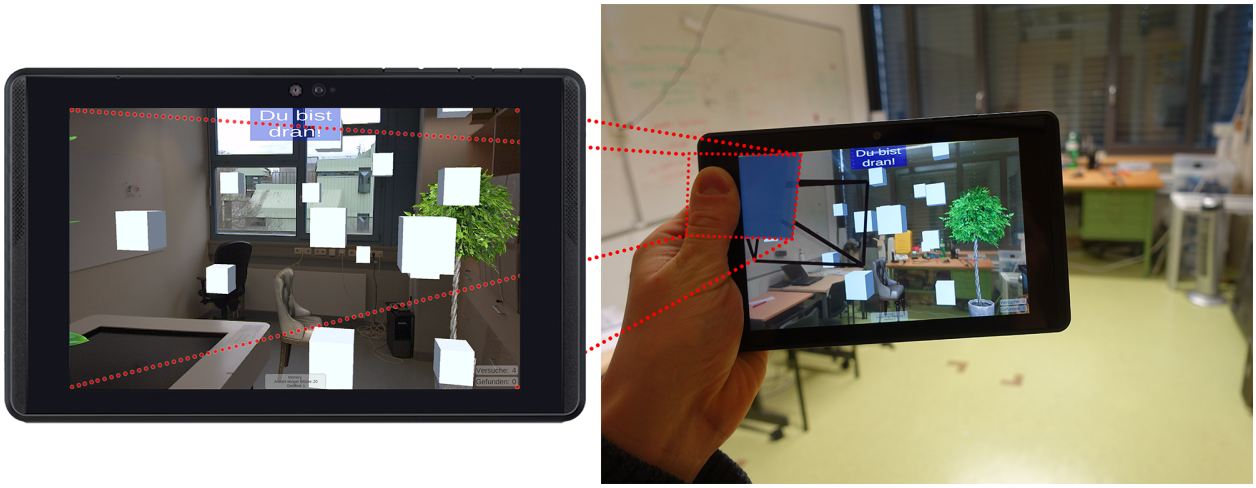


Figure 3.1: Final research prototype. Two collaborators are located in different rooms, thereby entering a semi-shared MR with a tango tablet. Both see the same virtual objects (plant, cubes). Additively, each player sees the tablet of his/her teammate (with field of view) as an abstract model (right, red dotted square).

As we intended to find out how spatial cues influence collaboration, participants completed each task twice: once with spatial cues displayed and once without. Figure 3.2 shows one room under both conditions.

Figure 3.2: Object identification task with (left) and without (right) virtual cues.

3.2 Technical Implementation

In this chapter, the complete technical approach and how we developed the necessary android application for the tablets is explained.

3.2.1 Google's Project Tango

In the scope of this research project, we used a handheld display with digital see-through as the medium to experience the MR. The *AR Information Browser* (Billinghurst et al., 2015) is especially suited in our case due to the natural movement and interaction possibilities it offers. To take a look around in the MR it is sufficient to move the display respectively. Interactions can be made by pressing digital buttons on the touch display.

Considering our application intent, Google's "Tango Tablets" (Fig. 3.3) seemed to provide a good basis for a convenient solution. The mid 2015 released tablets are equipped with three special modules, which we could use to implement the MREs: motion-tracking, area learning, and depth perception. They are designed to improve the device's capabilities with regard to visual measurement, indoor navigation, and augmented reality.
The "Project Tango Development Kit" comes with a SDK which provides the functionality to access the sensors. It also features a standalone exemplary Unity-Project which demonstrates the different capabilities of the tablet. Furthermore, an android application named "Project Tango Explorer" (Google Tango Explorer) can be used to create so-called "Area Description Files" (ADFs). Those files are needed to save spatial information in order to

Figure 3.3: Google's Project Tango Tablet Development Kit (Google Project Tango).

read them at another point of time and to re-localize the tablet in the physical environment. When the tablet is started, it needs to locate itself in the current environment. As there is no global/external tracking system similar to GPS, which tells the tablet where it is from the outside, the localizing is exclusively handled visually using the previously generated ADFs. The tablet basically scans the room again and checks if there are any matching patterns in the file.

Those ADFs have to be created initially for each room. If the room is altered largely, the ADF has to be extended or renewed. It is created by "scanning" the room using the mentioned "explorer" application. We scanned each room for about 15 minutes in different states (light on/off, minor object changes etc.) to improve the reliability of the file. This guarantees that the ADF works at all times of the day properly (we had some dyads later in the evening).
To establish working MREs, which implement the same virtual component in both rooms, we had to make sure that both tablets span their coordinate system in the same manner. For instance, it would be unfavorable if the origin is located at one bottom corner of the first room, while the same point is floating somewhere in the middle in the second environment. To prevent this, we aligned the virtual space equally in both rooms in such a way that floor and walls defined the origin of the two coordinate systems likewise in both environments. Thus, objects integrated and adjusted properly in one room (e.g., an object aligned on the floor) will automatically be displayed correctly in the second room as well, without adding any offsets to its position.

3.2.2 Technical Setup

We developed a central controlled study system. All actions, like the start of a program or the monitoring of the study, is controlled by the server. This simplifies the study procedure and makes it more controllable. The two tablets solely have to join the server. The different programs are then started by the study administrator on the server.
As illustrated in Figure 3.4, the tablets are connected via Wi-Fi to a local installed router. The server is registered in the same network. It runs the logging system (3.4 f), the game server (3.4 g) and a TeamSpeak 3 (TeamSpeak Systems) server (3.4 h) for communication purposes. We installed an action camera in each room to evaluate users' communication behavior afterwards.
In advance of the study, both tablets were connected to the local TeamSpeak server. The respective application on each device was started and connected to the central game server. Video and audio recording as well as the logging was started after the training rounds.

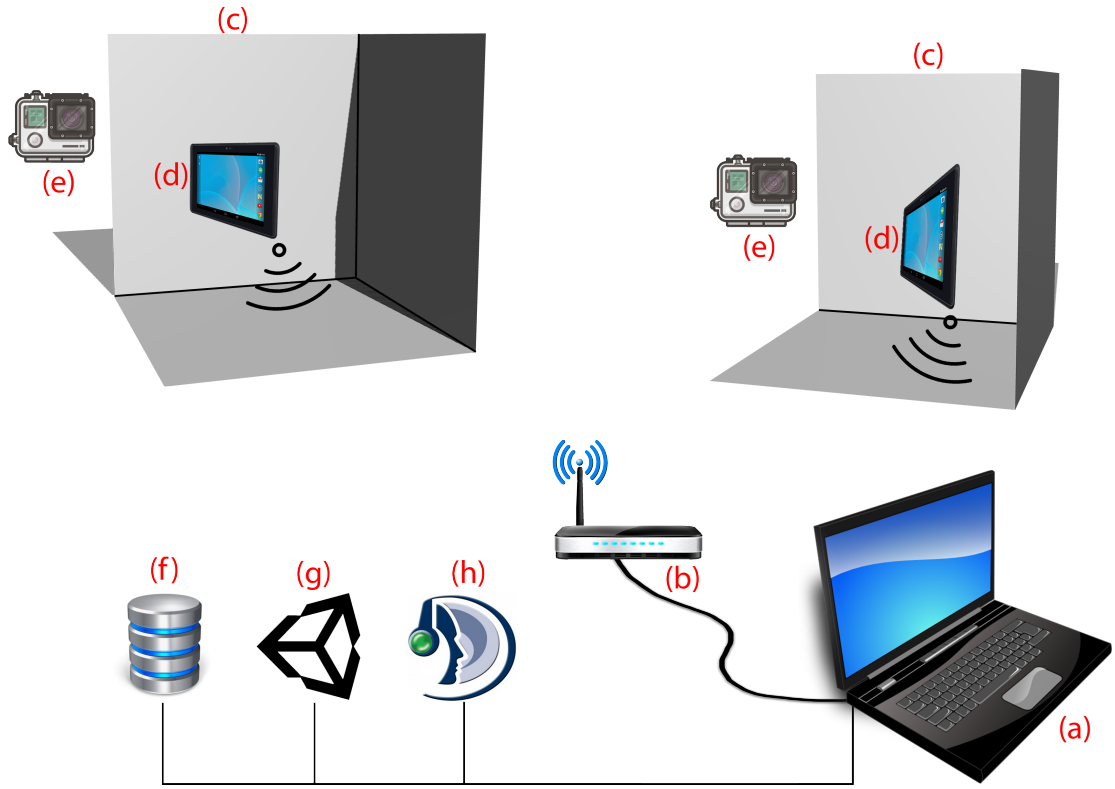


Figure 3.4: Setup of the study. (a) Centralized server controlling the game server (Unity (g)), the TeamSpeak server (h) and the logging system of the game (f). (b) Local WiFi router. (c) The two separate rooms in which the participants solved the tasks. (d) Each room has its own tablet. (e) Each room was video recorded using action cameras.

3.2.3 Implementation

Unity & Android Studio

In order to create the necessary test scenario we used the "Unity 5" framework in combination with the Project Tango SDK. Unity is a game development environment which can be used to develop applications for most common platforms. As our target platform was android, we installed Android Studio (Google Android Studio) for debugging purposes. Android Studio delivers functionality to attach a debugger to a process on an external device - in our case the tango tablets executing our study-application.

Development Process

The first step was to implement the basic functions such as see-through, localization, and the basic game functions. For the rudimentary basis we were able to use either core-scripts provided by Google itself or code snippets and complete scripts by Jens Müller (Müller et al., 2015). After that, we transformed the functionality into a networking-compatible system. Finally, we optimized the application using an iterative life cycle development model (Fig. 3.5), by always designing the next functional module, implementing it, and testing it. If the testing revealed that changes have to be made, we started a new iteration of the same process.

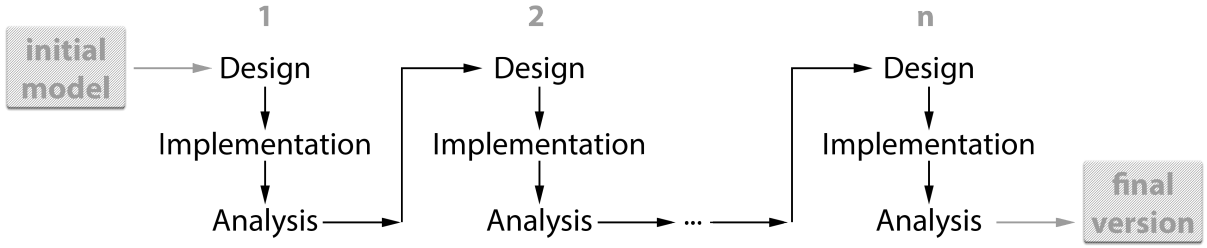


Figure 3.5: Iterative life cycle development model.

Implementation Model - Overview

As shown in a simplified manner in Figure 3.6, our prototype implements a basic server-client model. The server is supposed to run on an external computer while the clients run as android applications on the tablets and join the server. The entity `Player` not only contains a model for other players to be seen later in the MR, but also the logic for different modules of the client itself - e.g., the AR-see-through or the interaction-handler.
If an event is triggered by a client, for instance when a player taps on the touch screen,

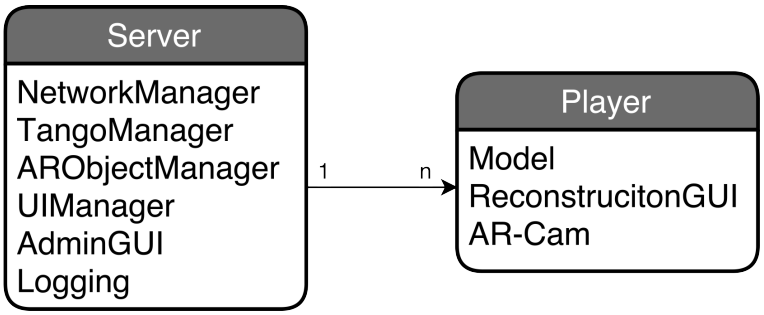


Figure 3.6: Key-functionalities of the entities `Server` and `Player`.

a `TouchEvent` (Fig. 3.7) is sent to the server. After that, the server evaluates the event and notifies all clients if necessary. In Unity, calls from the client to the server are called "commands", whereas server-side triggered updates on the clients are called "client remote procedure calls" (Client-RPCs). In order to evaluate a touch event, the server needs information about the player who triggered the event, the target which was touched, and the current game state. The last one indicates which program is currently running. Its consideration is important, as most interactions have distinct consequences in the respective tasks.

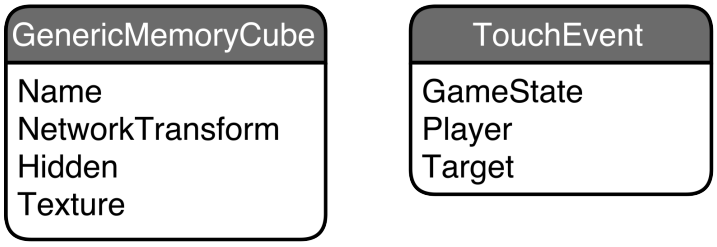


Figure 3.7: Key-functionalities of the entity `GenericMemoryCube` and the event `TouchEvent`.

On the server, a network-manager handles the communication between server and clients. Its functionality is completely provided in the Unity SDK. Besides, a so-called `TangoManager` is provided by the Tango SDK. It is required for controlling tango-internal core functionalities (e.g. control of the depth sensors).
We implemented an entity (class/object), which is responsible for all the virtual objects, which are spawned on the server, and their handling. The `ARObjectManager` basically contains all core functionalities relevant to our two tasks. For instance, it allows the examiner to spawn virtual cues or to start a game (memory/reconstruction). Another mentionable entity is the `UIManager`. It executes a script which facilitates additional textual displays like information boxes or error lock-screens. Moreover, in order to control the process of the

study conveniently, we created a simple GUI (`AdminGUI`), which the server administrator can use to start programs or to adjust initial parameters. Another very important module of the server is the `Logging` entity. It logs all information from the tablets, such as positions, orientations or interactions detailed to a xml file on the server. It automatically creates a folder- and file-hierarchy for each study distinguishing between the separate tasks and players.

In Unity, it is possible to create and use so-called "prefabs". They can be imagined as predefined forms or models from which objects can be spawned. We created such an entity for our memory cubes (`GenericMemoryCube`). It contains several attributes like a name and a texture. One attribute (`hidden`) indicates if the texture is currently displayed or not. If its value is changed, a command is triggered, which again triggers a Client-RPC in order to change the texture on the server and respectively on all clients (e.g., after a player clicked on the cube and opened it).

Technical Challenges

In the beginning we had difficulties implementing the networking mechanism. Figure 3.8 provides a simplified example of how networking is handled in Unity. Each component in the network (players, server) has a local structure of the complete game scenario containing all players, virtual objects, and game components. Each network component might have different active or inactive scripts. Some scripts can only be executed by the players, while others are executed by the server. For instance, if one player is using a script to move, a network call (Command) is started, telling the server that he moved. After that, the server starts a Client-RPC (Remote Procedure Call) which again tells all clients that this player moved. Subsequently, this player and its model is moved on all other players' tablets (analogous to the cube opening in Figure 3.8).

Sometimes some data packages get lost causing the system to be inconsistent. This involves the danger that a deactivated item is still active on one single player, but nowhere else (e.g., deleted memory cube). To prevent this, we tried to make the local scripts as independent as possible. This way, less information has to be transferred over the network reducing critical calls to a minimum. For example, we have an information box as shown in Figure 4.5(b). During the memory game, it displays how many cubes are left and how many are currently open. Usually, one would generate the information on the server and send it to the clients. This reduces calculation effort, but suffers the loss of reliability. As the computing time needed is not significant, we decided to outsource the calculations to each client itself. Therefore, every client updates the information every 0.5 seconds by autonomously counting the respective cubes. We used this principle wherever possible to relieve the server and the network, in order to improve consistency between all entities of the network.

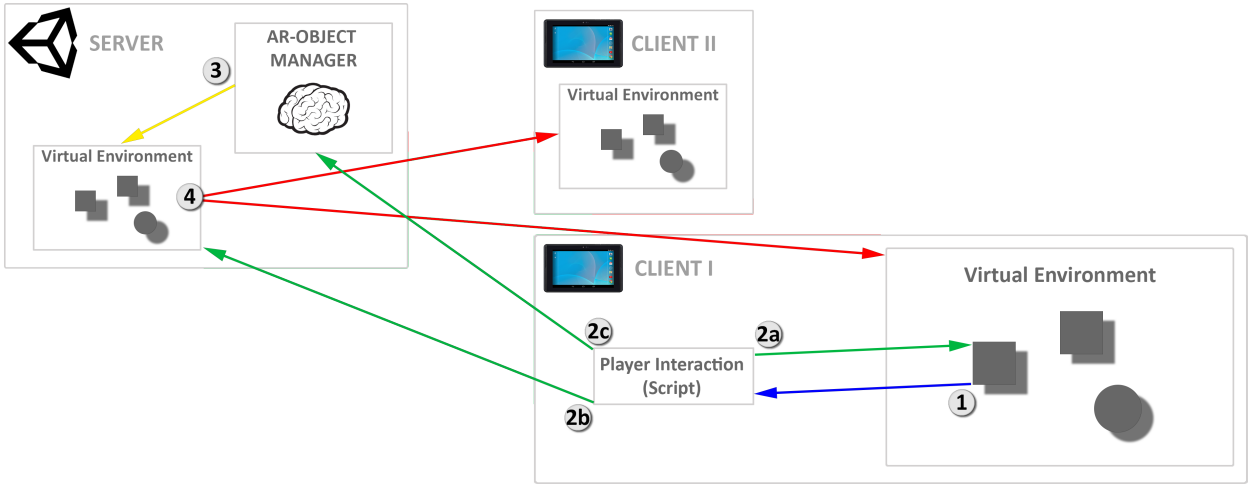


Figure 3.8: The detailed memory `OnTouch` event sequence. (1) Local cube throws event (it was touched). (2) Local script catches event and processes it by (a) changing the state of the local cube, (b) telling the server to change the state of its cube as well, and (c) sending a command to the AR-Object manager with the information about the changes. First, after (2b) is sent, the server automatically calls a Client-RPC, updating all Clients (4). At the same time, the AR-Object Manager processes the command (2c) and takes action if necessary (two cubes are open → close or delete them) by changing its own virtual environment (3). This again triggers (4) the change of all client virtual environments.

Another issue we faced, was that the whole study was at risk to be discarded in case a player was disconnected by any cause. The reason for this was that as soon as a player disconnected, its entity was deleted from the server, and the original state of the current game could not be restored. This means that if a player, for instance, sets two cubes in the reconstruction game, disconnects and rejoins, he would be able to set the same two cubes again. Similar issues occurred as well in comparable cases during the memory game. The player in turn disconnects and rejoins. Thereby, he automatically gets a new network ID. Based on this ID, the server decides whose turn it is - meaning none of the players can take action, as none of them has the proper ID. In case such a deadlock happened, and we would restart the server and the memory game, the players would already know half of the matching memory cubes. The results would be useless and the whole study would have to be repeated with other test persons or at least with different coordinate and texture sets. Therefore, we implemented a solution which allows players - even both at the same time - to leave and rejoin as often as they like to. A running program on the server keeps the play's status and does not abort just because a player is timed out due to network problems. Furthermore, the game is paused if a player disconnects in order to continuously obtain the same game conditions for all groups.

4 Study

The aim of our study was to find out how artificially added, virtual landmarks influence communication behavior, task load, user experience and social presence of users while mutually working on a cooperative task with a remote located partner. Particularly, we examined the following research question: *How do additional virtual objects shape collaboration in object identification tasks in physically separate mixed reality environments?* As previously mentioned, the development and conduction of this study was realized together with Matthias Miller. His focus was on the influence of virtual cues on spatial positioning tasks under the same environmental conditions.

4.1 Study Design

The study made use of an within-subjects design. Each dyad completed the two tasks under both conditions - with and without spatial cues (independent variable). This variable was counterbalanced for all 16 dyads.

4.1.1 Dependent Variables & Operationalization

The object of investigation was the influence of virtual cues in remote mixed reality environments on different aspects of collaboration. To begin with, it is essential to find characteristics which describe the collaboration process as accurate as possible. Only like that it is feasible to measure differences between the single iterations using the same participants with dissimilar preconditions in order to compare them afterwards and gain insight into the actual influence of the virtual cues. The generic term "collaborative work" is inexpressive, immeasurable, and mis-understandable when it stands for itself. So we operationalized the construct using four distinct dependent variables which describe "collaborative work" properly and which basically can be measured.

As shown before by Müller et al. (2015), spatial cues can have an influence on "communication behavior", "user task load" and "user experience" in co-located MREs. Therefore, we adapted these three measures for our purpose. As a forth dependent variable, we also included "perception of presence". Herewith it could be measured how present one perceives his partner in the remote MRE. In combination, these four variables are suited to convey a decent picture of the collaboration quality among the two tested persons in our study scenario

(Fig. 4.1). The communication behavior is measured by inspection of certain communication characteristics (video analysis). Information about the subjectively perceived task load and presence are gathered using questionnaires (NASA TLX: Appendix A.1 TLX Questionnaire, and extended Temple Presence Inventory (TPI): Appendix A.2 TPI Questionnaire). The user experience is measured by means of user statements in the final interview.

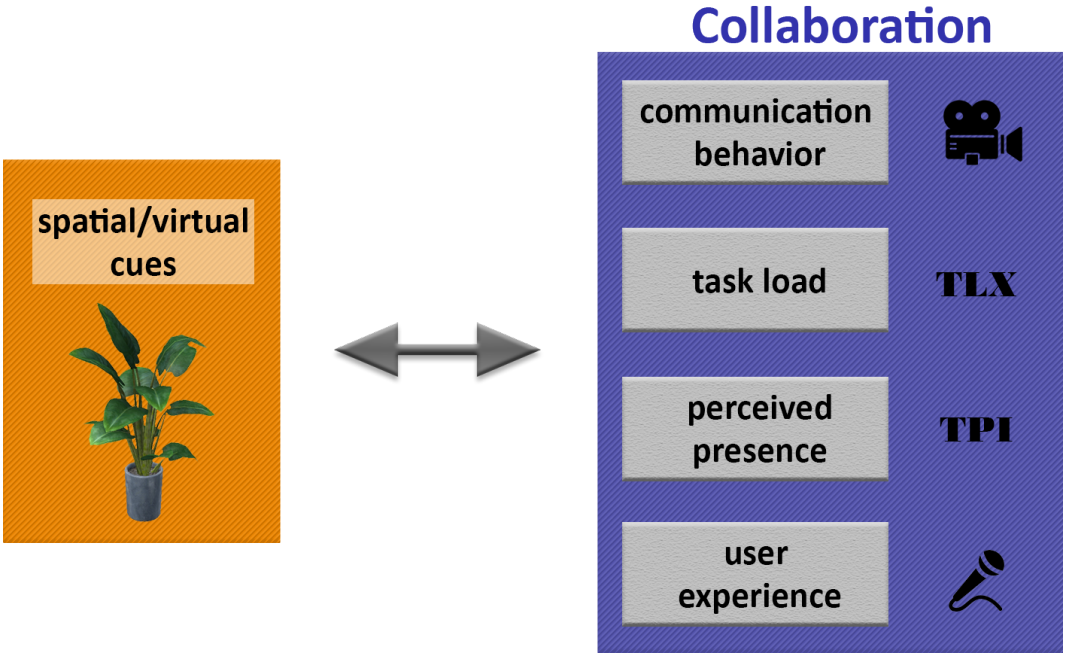


Figure 4.1: Our prototype covers all four dependent variables. "Communication behavior" is analyzed by use of the video footage from the experiments. "Task load" and "perceived presence" are collected via questionnaires (and data logs). "User experience" is measured by means of a semi-structured interview in the end.

Note: for our study we used a slight modification of the TPI questionnaire (complete form is attached in the appendix: Appendix A.2 - TPI Questionnaire). This variation concerns the inclusion of only two dimensions of the original TPI, as they were the most fitting ones in terms of the purpose of our study. First, two novel questions (not contained in the original version of the TPI) were asked, which geared directly towards our field of investigation (*How were the possibilities to coordinate actions between you and your partner?* and *How were the possibilities to communicate spatial information with your partner?*). Then, two standardized dimensions of the TPI were used. The dimension 'social presence - actor within medium (parasocial interaction)' included 7 questions concerning participants' perception. These should be answered on rating scales raging from 1 to 7 (e.g., "*How often did you have the sensation that people you saw/heard could also see/hear you?*"; *never*=1 to *always*=7). Next, the dimension 'social richness' should be assessed on the basis of seven aspects which could be used to describe their media experience (e.g., impersonal=1 to personal=7).

One possible way to examine "collaborative work" by means of the previously named dependent variables is through observation of the test persons while they execute predefined tasks. Therefore, subjects were asked to perform a collaborative object identification task and an object positioning task - each one once with and once without virtual cues displayed. Participants were urged to complete the same tasks twice, under both conditions. This procedure guarantees that both rounds pose similar challenges and prevents differences arising from divergent difficulty levels (counterbalanced design).

We used a number of different techniques to gather enough information to cover all our dependent variables sufficiently in order to receive convincing and meaningful results.
To measure our first dependent variable "communication behavior", we recorded the complete conversation during the tasks in order to analyze it manually later on. Hereby, the focus was on the usage of spatial expressions. Among others, interesting questions were whether the frame of reference is determined or influenced by the display of virtual cues. Or if virtual cues have an impact on the amount of used deictic speech.
The second variable "task load" can be measured by means of a questionnaire after the task, asking the participant how he or she perceived the overall work load during the task (NASA TLX questionnaire (Hart and Staveland, 1988)). Moreover, the log-data of the actual procedure of the game is considered as a source for this dependent variable. It can be analyzed manually by taking a closer look at the players' interactions, occurred events (e.g., memory cube openings) and meta data (e.g., how much time a group required to solve the task and how many attempts to find all matching pairs were warranted).
Characteristics of the third variable - "perception of presence" - can also be obtained using a questionnaire. As the name of the variable already predicts, it measures the test person's subjective opinion on how physically close or remote the other person felt. Therefore, this variable opens up the possibility to reveal if it is significantly easier for a person to "step in" the shared virtual environment and navigate in the room when spatial cues are present. Usually, orientation and navigation within a group of people pose no serious difficulties in case all its members are situated in the same room, as people are very much used to this (Fussell et al., 2000). According to Fussell et al. (2000), the same applies to collaboration, and it would be ideal with respect to efficiency, communication flow and so on, if both persons were actually in the same room (or at least if they would have this feeling). This again is due to the higher level of their common conversational grounding as described by Clark and Brennan (1991). Their grounding is better, when both collaborators are co-located, see each other and are able to speak with each other - due to more possible reference instruments (visibility, audibility, simultaneity,...). Therefore, if the spatial cues could help people to feel more present in the mixed reality (with their remote partner) it would implicitly show that the cues establish a better common grounding. In this case, virtual cues would help collaborators to overcome their remoteness by improving the conversation and collaboration abilities of the group.

Last but not least, data to evaluate the overall "user experience" was gathered by use of a semi-structured interview at the end of the experiment. This is a convenient way to capture users' preferences and reasons for their impressions, personal tendencies, and opinions - findings the data analysis could probably not reveal. A further motive for evaluating this variable is that the individual user experience can influence the collaboration flow by means of individual effectiveness, productivity, and decision making abilities as shown by Bubaš (2001).

Round	Game	`Cues`	`CoordinateSetUsed`	`TextureSetUsed`
1	Memory	Yes	1	2
2	Reconstruction	Yes	1	2
3	Memory	No	2	1
4	Reconstruction	No	2	1

Table 4.1: Exemplary study setting containing two tuples of tasks. For each tuple the settings are the same and are inversed for the second tuple of tasks.

As exemplary shown in tabular 4.1, each task was played two times - once with spatial cues and once without. The variable `Cues` with the characteristics `Yes` and `No` is our independent variable. Our dependent variables are: user task load (NASA TLX), communication behavior (audio and video analysis), subjective perceived presence (TPI) and user experiences (semi-structured interview). All variables were counterbalanced for all 16 dyads (display of cues, used coordinate set, and used texture set).

The two tasks belong together, respectively (i.e., memory and reconstruction; see Tab. 4.1 1+2 and 3+4), as both use the same settings. We used different spawn coordinates (`CoordinateSetUsed`) and textures (`TextureSetUsed`) for the memory cubes during each pair of tasks. For both parameters - `CoordinateSetUsed` and `TextureSetUsed` - two states were possible as we prepared exactly two sets each (e.g., for the `TextureSetUsed` 10 Wingdings textures each - see Fig. 4.4). In order to minimize possible distortions through sequential effects (carry-over-effects), like training or fatigue effects, we randomly organized the order of the initial parameters (coordinates, textures, and spatial cues). We cycled through all variations, resulting in 8 (2^3) different start settings (start parameters; pertained for the first half, i.e., memory + reconstruction). Then, each group had the respective inverse settings in the two tasks of the second half (change of condition (cues), coordinate set and texture set). In addition to that, we started the study with a training round, using different textures (4.4c) and coordinates in order to give research participants the possibility to get familiar with the tasks and to ensure that all participants understood the game sufficiently.

4.1.2 Study Task

As basis for our physical layout serves a setup of remote MREs (our intentions require a scenario which provides the optional overlay of virtual cues - our independent variable). In order to establish a mixed reality, a medium is required to make the virtual component of the MR visible for participants. We chose Google's Project Tango tablets (Google Project Tango, 2016) as they feature all technical requirements. The virtual component of the MR is rendered into the see-through video stream with proper scaling and spatial distortion, so both worlds blend, conveying the impression that the virtual objects are actually located in the room.

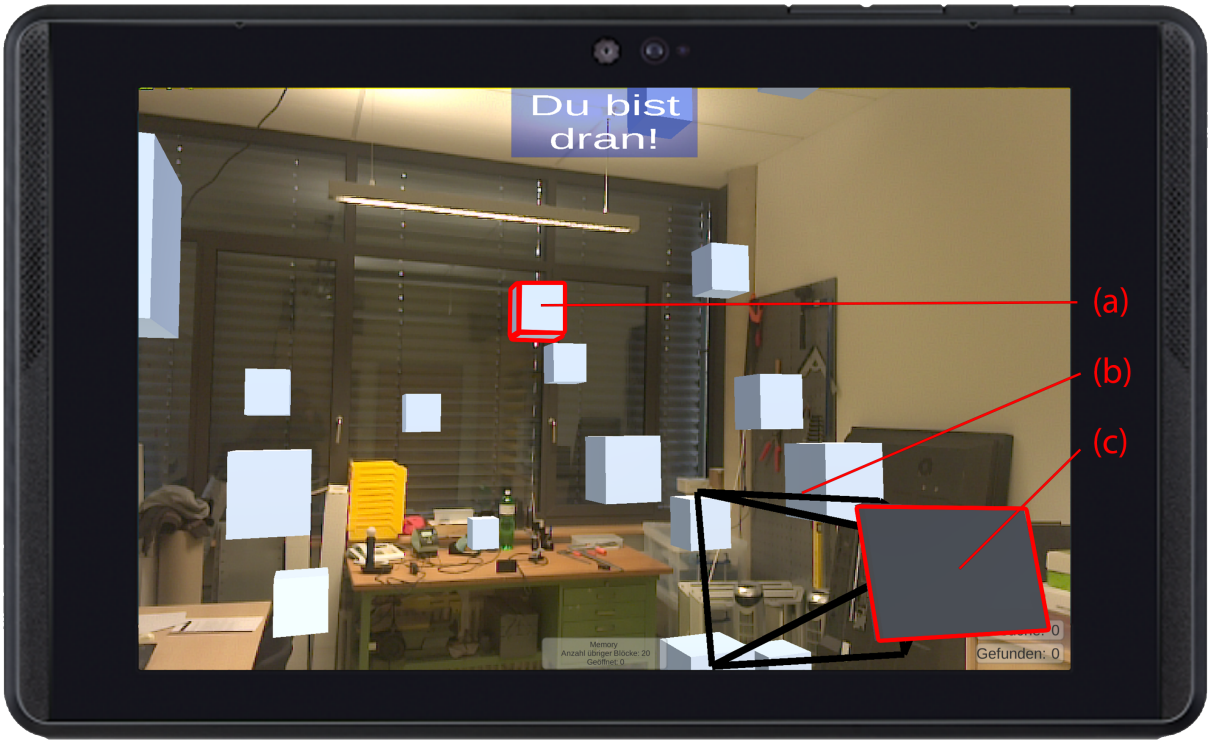


Figure 4.2: (a) Memory cube. (b) Teammate's field of view. (c) Teammate's tablet.

Once our goals were specified, we faced the task to develop a scenario which was capable of measuring all employed dependent variables (Fig. 4.1). Virtual cues can have a large influence on the collaboration in the same physical environment (co-located), as recently shown by Müller et al. (2015). Our intention was to extend this finding to scenarios with remote instead of co-located collaboration. Thus, the main point of interest was to investigate the influence of spatial cues on collaboration. Additionally, we tried to examine if spatial cues are (equally) important or even gain in importance, in case users are situated in distinct physical environments compared to them being co-located (using the results by Müller et al. (2015)).

To infer relationships between and to compare our study to the one conducted by Müller et al. (2015), we built up the test environment very similar with only minor changes. Each participant was located in a different room. Both had a *Google Project Tango Tablet* (Google Project Tango, 2016), with which they had to solve several tasks together with the partner in the other room. They were able to speak to each other via the tablets. Furthermore, they could see game components (Fig. 4.2a) and the tablet of their partner as an abstract model (Fig. 4.2b, c) through their tablets as parts of the MR.

There were two different kinds of tasks, which they were supposed to do. First, they were encouraged to play a modified 3D version of the memory card game. As shown in Fig. 4.2, players had to find matching pairs of memory cubes by opening them (first task). By the end of this game, all memory cubes were gone. After that, they were asked to place the cubes of the memory game at the respective location where they had been before (second task). More detailed information about the two tasks is provided in section 4.1.2.

For the work in hand, the first task is of particular relevance, as participants have to solve an object identification task in remote collaboration. In doing so, they communicate in order to find matching pairs efficiently. Afterwards, all variables can be measured by either evaluating the audio, video, or log recordings or through consideration of their answers in the interview or the questionnaires. The semi-structured interview offers the opportunity to dig even deeper and reveal opinions and attitudes of participants that neither the log data nor the questionnaires could display. In the second part, participants were not solving an object identification task directly. Nevertheless, their overall score might have been influenced by the previous task (memory). Moreover, on its basis Matthias Miller investigated the cues' impact on object positioning tasks.

Figure 4.3: Interface of the object positioning task. Button for finishing level in red.

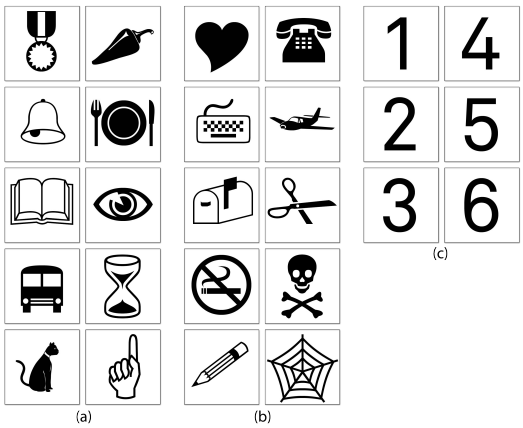


Figure 4.4: Wingdings textures used for memory cubes. (a) Set 1. (b) Set 2. (c) Training set.

Object Identification Task (Memory)

Each dyad completed two different types of tasks. The first one was a 3D version of the memory card game, in which players had to open AR-cubes in order to find matches. Those memory cubes, which represent memory cards, were distributed in the MRE. The cubes, which were "floating" in both players' common MRE, could be opened by clicking on them using the touch screen of the tablet. Once a player had clicked on a cube, a texture was rendered over the cube, displaying its Wingdings font (Fig. 4.4). If two cubes were open at the same time and their textures matched, they were deleted from the MRE. If not, their texture was hidden again. The game terminated as soon as all pairs of cubes were deleted. The player took turns - one player after the other had to open one cube at a time. An information box at the top of the screen indicated whose turn it was (see Fig. 4.5a). This was an additional constraint to make sure that the attendants exchange their memories and try to lead each other to the - in their opinion - right decisions. It was introduced to stimulate communication and to promote the overall collaboration component in this task.

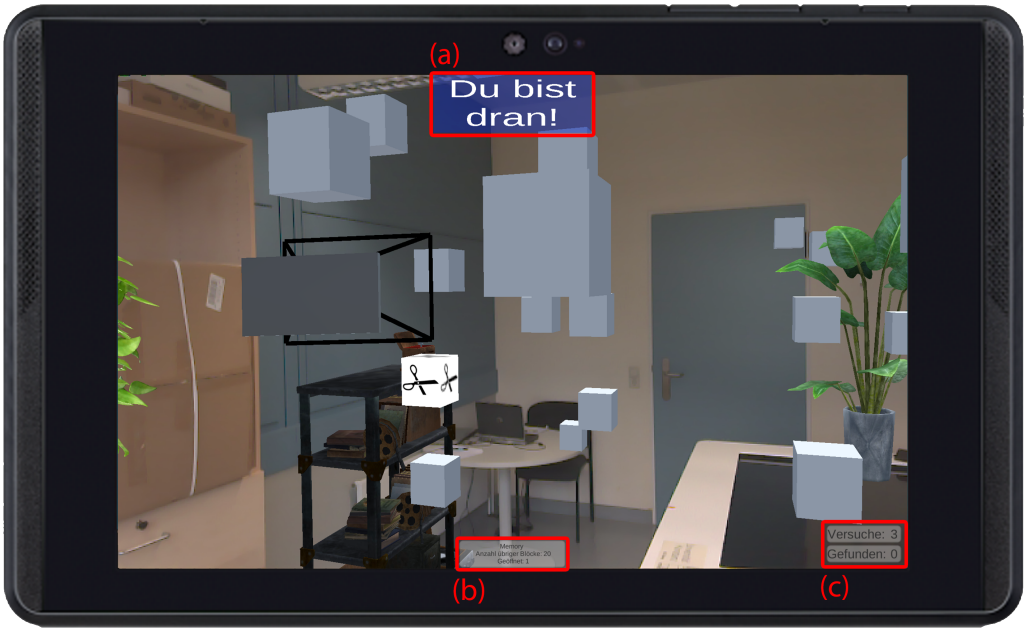


Figure 4.5: Interface. Screen-shot during object identification task (4.1.2). (a) Memory only; Info-box indicating whose turn it is. (b) Current game info. (c) Game stats / Score.

Figure 4.5 shows how the interface for this task looks like. The information box at the top center of the screen (4.5a) is solely displayed in the memory task. It indicates whose player's turn it is to open the next cube. The GUI embeds two more information displays. Both are located at the bottom of the screen (4.5b, c). The center one (b) holds information about the game status and the environment, for example how many cubes are currently

displayed and how many are open. The second one (c) in the right bottom corner displays the current score of the team. This is set up in order to motivate the team to get the task done efficiently.

Object Positioning Task (Reconstruction)

The second type of task, which the dyads had to solve, is the "reconstruction" task. Within this task, players had to reconstruct the constellation of memory cubes of the previous memory task by placing the cubes. The GUI for this task is shown in Figure 4.6. Each player had 10 different cubes (4.6c). One from each pair of the previous task.
Each player carried a semi-transparent cube in front of him, which indicates where the cube was spawned after a button was pressed. Once one of the buttons on the sides of the screen was pressed, the cube with the respective texture was spawned and the button was marked as inactive (gray overlay - e.g., button "telephone" in Fig. 4.6c). In order to reactivate the button and to set the cube again, the player first had to destroy the existing, previously set cube in the MR by tapping on it (touch display). If a player changed his mind or accidentally misplaced a cube, he could simply click on the respective cube and it was taken back in his inventory. The game terminated when both players had set all 10 cubes and had pressed the "Finish Level!" button (Fig. 4.3).

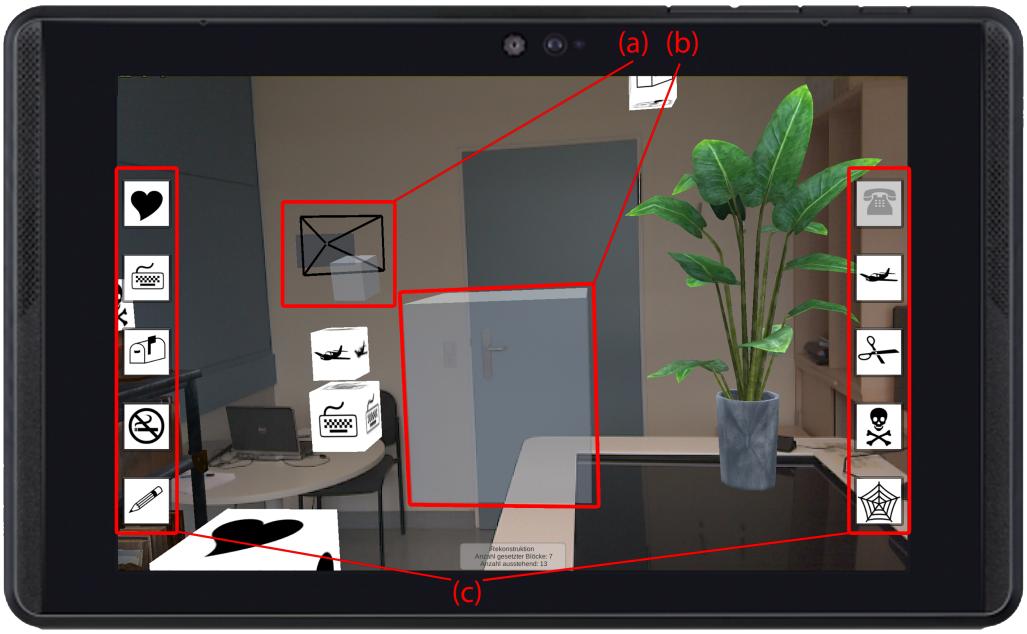


Figure 4.6: Interface object positioning task. (a) Teammate with transparent cube. (b) Own reconstruction cube indicating where new cubes are spawned after pressing a button. (c) Buttons - gray if inactive.

The semi-transparent cubes were floating 60cm in front of each player, having the same size as the memory cubes (15×15×15cm). They were oriented exactly like the original cubes and therefore did not rotate or tilt with the player.
While participants could stay at one point of the room during the memory task, during the reconstruction task, they had to walk around in order to place the cubes where they think they have been before.

4.2 Experiment

The execution of the study took place from January 12 to January 15, 2016. One week later we invited two more dyads to participate in the study as replacement for two inapplicable datasets from the week before.

4.2.1 Study Procedure

Our study was structured as follows: After a short welcome introduction, the two research participants signed a declaration of consent, so that we were allowed to use the video, audio, and log materials afterwards. Subsequently, they filled out a demographic questionnaire. After that, we started with a training unit to ensure none of the two conditions (with and without spatial cues) is privileged by simply being the second one proceeded.
When the participants did not have any further questions, we started the first game followed by two questionnaires. This process was repeated four times until two pairs of each a memory and reconstruction task were completed. Directly after that, we conducted an interview with both participants together at the same time. Table 4.2 provides an overview of the whole process.
During the different tasks, each participant was recorded with a fish-eye action camera to cover each room completely. Besides, we recorded their communication via the tablets, using TeamSpeaks' internal recording system. Moreover, we logged all information from the tablets (positions, matching states (e.g., initializing, valid etc.), and user interaction events).

4.2.2 Apparatus & Study Environment

The two rooms we used had about the same measurements (5.5x3x2.4 m). However, they were oriented reversely, as one of the rooms was on the western side of the hallway, while the other one was on the eastern side. Except for the orientation, the architectural layout of the rooms differed little from each other, as only the doors did not match their location exactly (still same side of the room). On the contrary, their interior was very different. One room, in the following referred to as room 907, was quite a mess and therefore exhibited many real features/physical landmarks, whereas the other room (923) was rather clean and sterile only containing one large touch table, a smaller table, and a chair.
We adjusted the virtual environment at the side of the door (right corner). That means, objects which were close to the windows in one room were also nearby the window in the other room. Furthermore, the virtual overlay had the size of the real room.

Nr.	Description	
1	Welcome	
2	Declaration of consent	
3	Demographic questionnaire	
4	Introduction and Training	
5	**Memory**	condition A
6	TLX	
7	TPI	
8	**Reconstruction**	
9	TLX	
10	TPI	
11	**Memory**	condition B
12	TLX	
13	TPI	
14	**Reconstruction**	
15	TLX	
16	TPI	
17	Interview	
18	Compensation	

Table 4.2: Procedure of the study. Both tasks are conducted twice. Conditions change between the two runs (one time with cues, the other time without cues).

Figure 4.7: Virtual objects, which were inserted into the MRE. Ceiling lamp, plants, bookshelf, chair (Unity 3D Asset Store 2016).

As our physical environments were a little smaller compared to the original study of Müller et al. (2015), we also adjusted the cubes' size from 0.25m to 0.15m. We distributed 20 cubes all over the room, using the natural boundaries of the rooms as constraint and making sure that they did not intersect with a real object in one of the rooms.
In our case, the medium to facilitate the blending of the virtual component into the real environment was a pair of Google Tango tablets. Their 7.02" display (323 ppi) allowed users to view the mixed reality.

We used five different virtual objects as additional visual landmarks displayed during the conditions with cues. The 3D models were provided by the "Unity Asset Store" (Unity Asset Store 2016). As shown in Figure 4.7, we used two plants, a chair, a chandelier-lamp, and a bookshelf. They were distributed all over the room (Figure 4.8). We only considered positions which were free of physical objects in both rooms. One item (plant) was displayed in about 90cm height. It was one of the few positions at which both environments had a free surface on the same height (desk (907), touch table (923)).

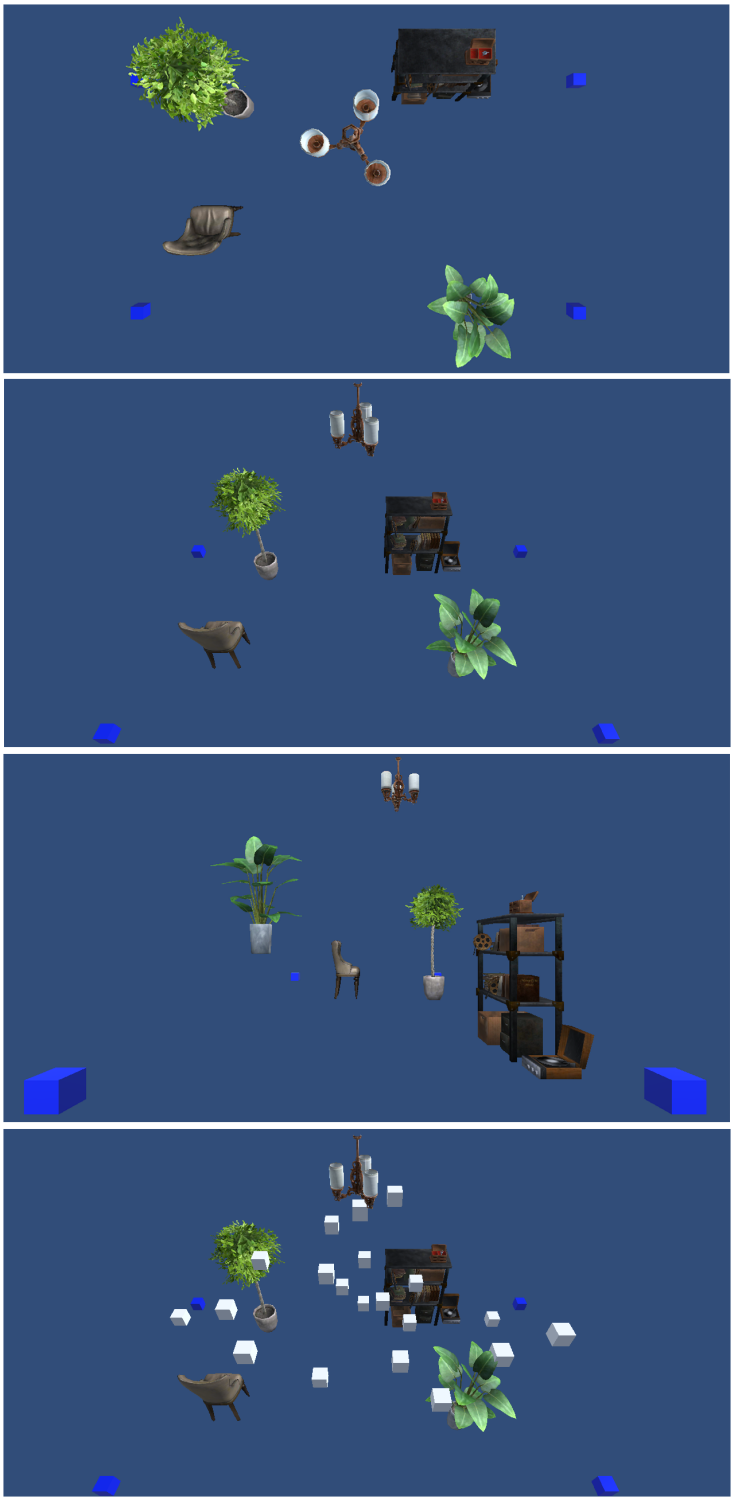


Figure 4.8: Overview - setup of the virtual room. In-game non-visible border cubes in blue marking the corners of the real rooms. Bottom: with memory cubes.

4.2.3 Participants

We conducted the study with 19 dyads in total. Before the official study, we performed a test run. As several minor things were changed after this test run we did not include it in the analysis. Of the remaining 18 dyads, 16 were used as basis for performing statistical tests. Technical problems prevented the inclusion of the two other dyads (9 and 13).

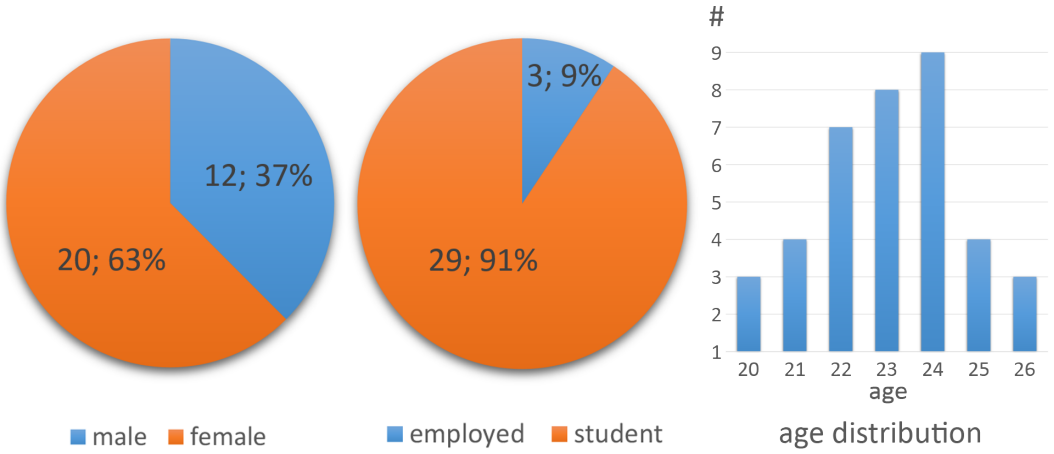


Figure 4.9: Demographic overview. Left: gender distribution. Center: occupation of participants. Right: age distribution.

Of the 32 participants 12 were male and 20 were female. They were between 20 and 26 years old (M=22.94, SD=1.68). About 91% were students, the others pursued a profession (see Figure 4.9 for a quick overview). About a third uses a tablet on a regular basis (on average for 2.3 years by now - more detailed fragmentation in Figure 4.10).
Four participants stated that they already had contact with the subject matter of augmented reality. The experiences mentioned were in the context of AR smart phone applications, "Oculus Rift"[1] applications, university presentations, and different studies they took part in. Most of the dyads did know each other before attending the study. Only three groups newly met. Two test persons knew the physical environments of both rooms before participating in the study.

[1]Oculus Rift: virtual reality head-mounted display developed by Oculus VR (2016).

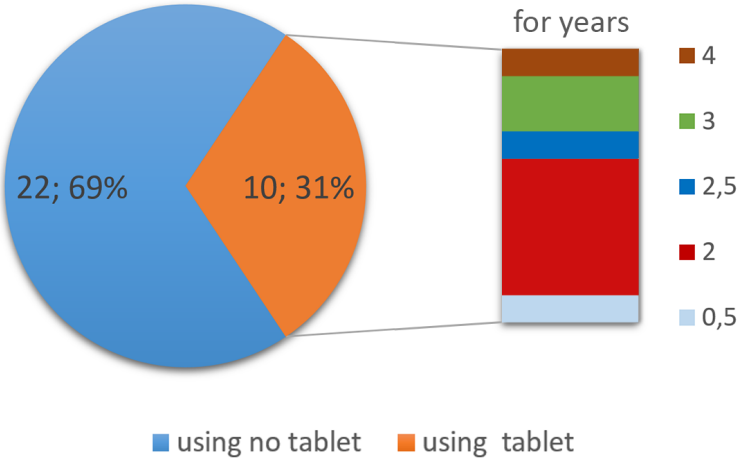


Figure 4.10: Distribution of tablet interaction experience. Right: partition of experienced users - duration of tablet usage.

4.2.4 Evaluation Approach

In this section, the approach of how the gathered data was evaluated will be explained. The outcomes are then presented in the next section (4.3 Results).

During the process of the study we recorded video and audio. We also logged the complete interaction with the tablets. After each task, participants completed two different questionnaires (TLX, TPI). In the end we conducted a semi-structured interview. Therefore, we had four main sources to evaluate. Video and audio files were analyzed manually and all occurred references were categorized. We applied statistical techniques to evaluate the two questionnaires. Last but not least, we preprocessed and visualized the data from the log files using SPSS and R (IBM Corp, 2015; R Core Team, 2015).

Video Analysis

In order to analyze the recordings of the study, we first merged and synced the two videos (the respective one from each room) and the TeamSpeak recording, we received from the tablets (Fig. 4.11). In the subsequent analysis process we categorized all occurring expressions and noted down further interesting characteristics. The categories were determined in advance by analyzing the transcripts of the whole study and a smaller sample of two study runs using the video footage.

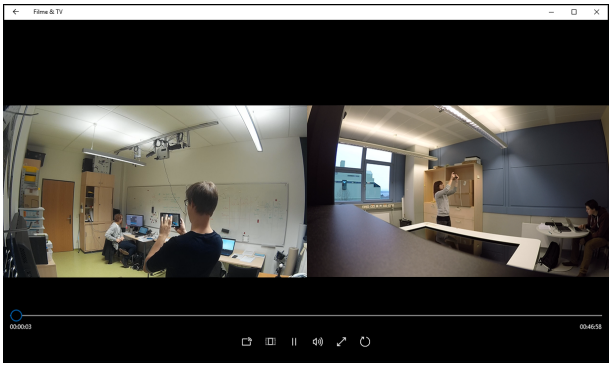


Figure 4.11: Merged and synchronized videos for manual analysis.

We decided to perform a very detailed categorization as the transformation to a less detailed level would be easily possible afterwards (see Figure 4.12). In our case, the references can first be grouped by their absolute, relative, and deictic property. Examples would be “at the table” (absolute), “above the table” (relative), and “over there” (deictic). Deictic speech covers all expressions that cannot be completely understood without additional information (“here”, “there”). After that, the expressions are classified more detailed by the point from where the reference originates. I.e., “left of the plant” would be a relative reference based on the position of the plant and therefore be classified as “relative to virtual cue”. We created a class for each combination (absolute/relative + cube/person/...). In the end, we had 19 different classes which could be grouped as desired. Like this, one could group the occurrences of self-references - regardless if relative or absolute - and still extract all absolute references at a later time. The complete evaluation form with examples for each class is attached in the appendix (Appendix B.1 - Evaluation Video & Audio Footage).

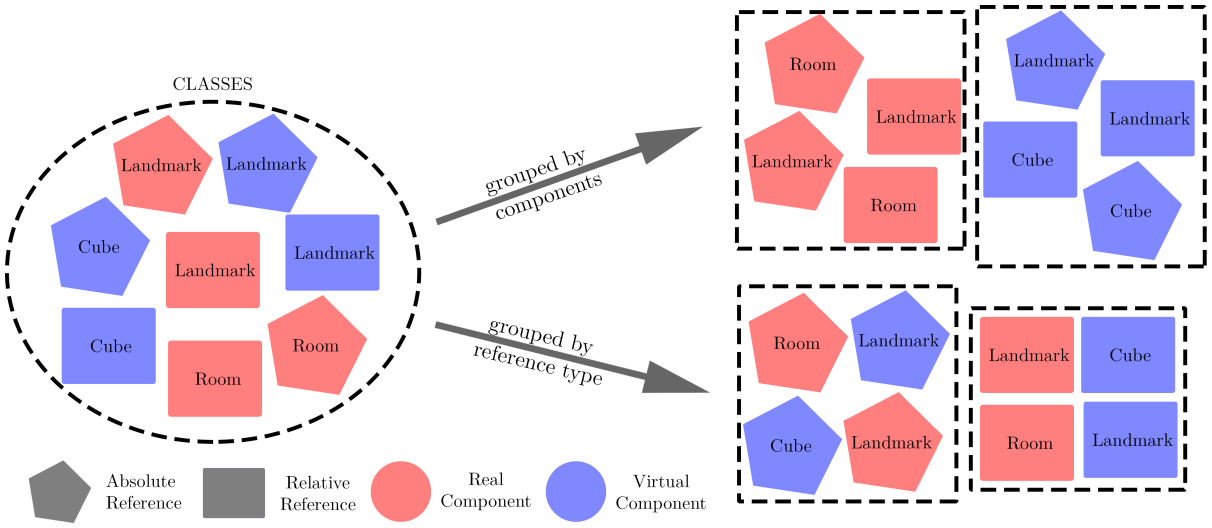


Figure 4.12: Detailed classification for subsequent dynamic grouping by desired characteristics (e.g., by type of reference (absolute/relative) or by type of object (virtual/real)).

In order to categorize as uniformly as possible, the two of us made a test run on one random dyad, categorizing its references and discussing misapprehensions and disagreements. In the end we had the same grounding for the process. Next, each of us categorized eight passes and additionally three in order to check the gained values for deviations. Within each dyad we classified the spatial expressions separately - not only for each task, but also for each room. By this means, we kept open the possibility to distinguish between the rooms afterwards and, if necessary, to examine differences due to the distinct nature of each room (one room was very plain whereas the other featured many items / real spatial cues).

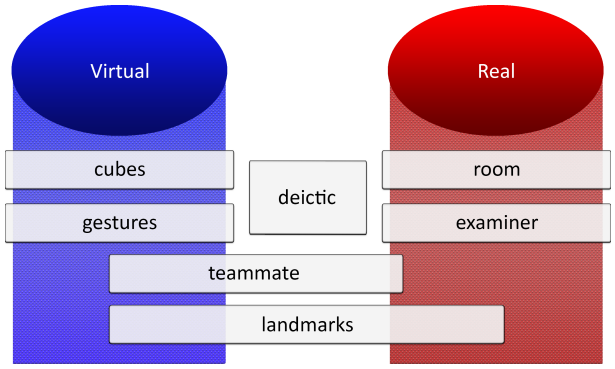


Figure 4.13: Categorization by type of object. Super-classes: virtual objects, real objects.

Figure 4.13 illustrates a model of all possible references for evaluation purposes. As our research question is directed towards how virtual cues influence collaboration, we grouped the classes by the type of the referenced objects. For our purpose, it is especially relevant whether these are components of the real or of the virtual environment (memory cubes or spatial cues).
Deictic expressions are the only 'references' which have neither a virtual nor a real basis point from which they originate. Gestures can essentially be

categorized as virtual references as the actual gesture is performed using the virtual model or the reconstruction cube. In very rare cases, test persons actually tried to point somewhere using their hands, recognizing promptly that their co-worker is not able to see it. In such cases, we noted this in a separate section titled "additional characteristics". As one can see in Figure 4.13, references with regard to the teammate are located more in the virtual section than in the real one. This is due to the fact that if references in relation to the teammate were made, they actually took reference to his/her model. Nonetheless, it could happen that a participant references the position of the other without seeing him (i.e., when he references his actual position).

After the footage of all dyads was categorized, we applied statistical methods to identify patterns and to determine trends. As mentioned before, we grouped several classes to determine relationships between their super-classes (e.g., real vs. virtual, absolute vs. relative or room 907 vs. room 923).

Questionnaires' Evaluation

The NASA TLX questionnaire consists of six rating scales (Figure 4.14 bottom). In combination, they constitute the participant's subjective overall task load. Nevertheless, they can also be considered separately (Hart and Staveland, 1988). We used an alpha level of .05 for all statistical tests.

To exploit all possible correlations, we proceeded as shown in Figure 4.14 (top). For each task, we evaluated the data by cues (with/without cues) and by rooms (907/923). I.e., we evaluated if changes of the respective conditions (cues → no cues; room 907 → room 923) caused significant changes on any of the six TLX scales or on the overall subjective work load (combination of all six scales).

To begin with, we extracted the descriptive information for all paths. Then, we applied the Shapiro-Wilk normality test (Figure 4.15 A) to determine, which of the TLX rating scales (Fig. 4.14 bottom) were normally distributed in order to choose the appropriate statistical method (parametric or non-parametric). Figure 4.15 provides an overview of the different applied statistical methods.
We differentiated between the classification by cues

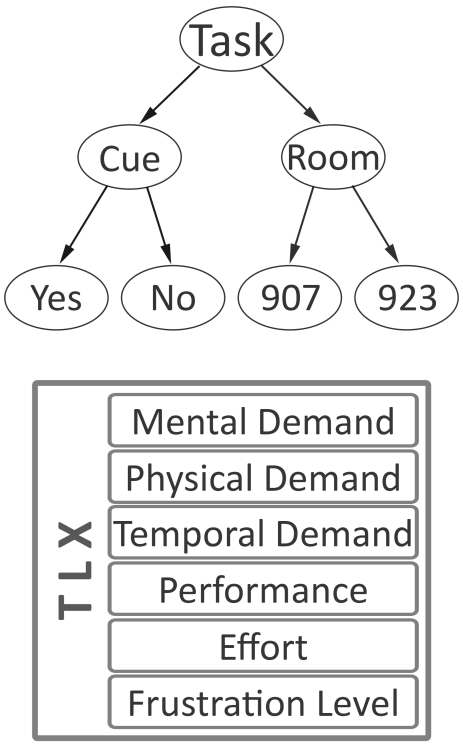


Figure 4.14: NASA TLX. Top: for each task, the results can be ordered by cues (with/without cues) or by rooms (907/923). Bottom: six TLX rating scales.

and by rooms for further processing. Participants completed the tasks both with and without cues (within design), but each participant stayed in one and the same room during the whole experiment (between design). Therefore, we applied different non-parametric procedures (Fig. 4.15 C). When grouped by cues, the Wilcoxon signed-rank test was utilized (Fig. 4.15 D,G). However, when analyzed by rooms, the Mann-Whitney test was employed (Fig. 4.15 E,H). For parametric data, t-tests were used (Fig. 4.15 B,F).

The structure of our approach for the evaluation of the extended TPI was similar. Here, we also distinguished by cues and separately by rooms. The extended TPI comprises 16 and not only six (TLX) rating scales. The same statistical methods were applied.

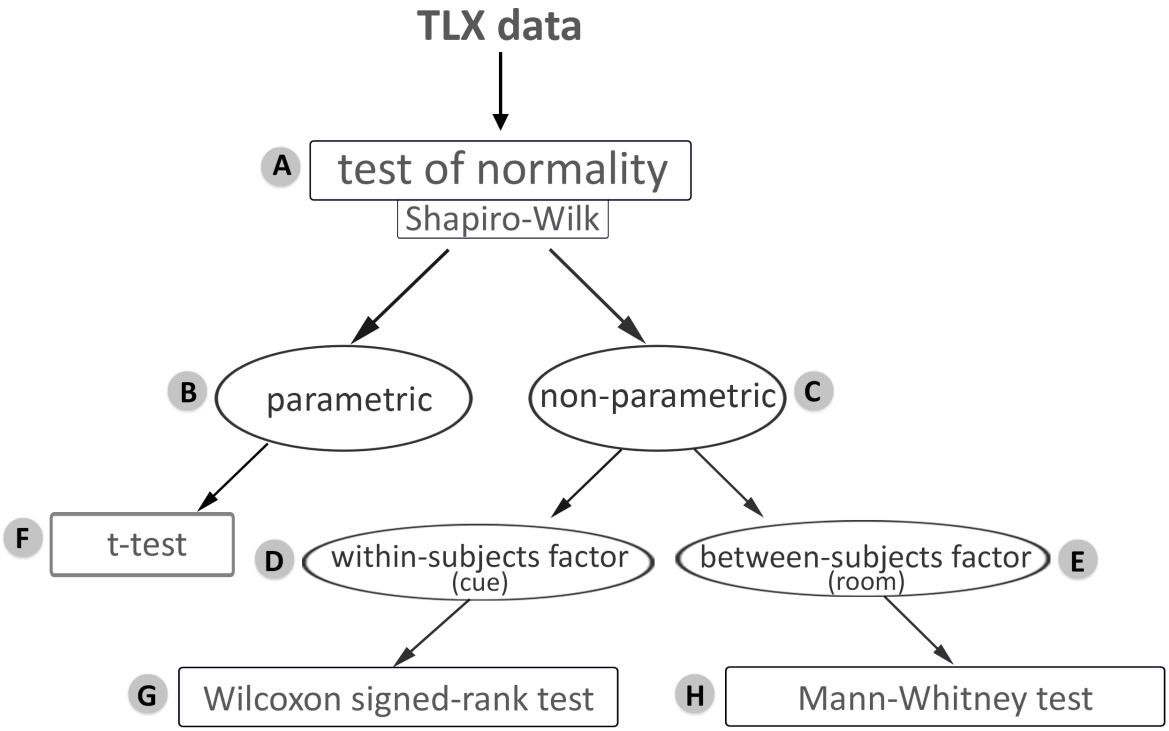


Figure 4.15: The statistical methods applied on the TLX data. First, each TLX rating scale (e.g., mental demand, effort,...) was tested for normal distribution (A). Normally distributed (B) rating scales were tested with the t-test (F) on significance. For the non-parametric data one has to differentiate between within-subjects factors (when grouping by cues was applied; D) and between-subjects factors (when grouping by rooms was applied; E). For within-subjects factors we applied the Wilcoxon signed-ranks test (G), whereas we used the Mann-Whitney test (H) for between-subjects factors to reveal if the differences between groups on the respective rating scales were significant.

Interview & Log Files

The goal of the interview evaluation was to determine what the overall user experience was and how it was influenced by the virtual cues. In our notes of the interviews we manually searched for patterns and repeating opinions. In the end we tried to draw some basic statements from the interview. Particularly, we were interested in the ones which might affect collaboration.
The log file analysis aims to serve supportive information by showing marginal differences caused by the two study conditions. I.e., differences in areas which are not directly targeted by our design such as the task completion time. Nevertheless, they can still be taken into consideration with regard to any of our dependent variables or just as further descriptors for the quality of the collaboration.
As mentioned before, the server logged the complete tasks including all players positions, orientations, and interactions. For each tested dyad the data was stored in a separate folder hierarchy. In order to process all the data and to import it in SPSS we transformed it from XML to CSV, using a construct of regular expressions. After that, we applied statistical methods on the data (normality tests, significance tests).
For instance, we intended to find out if spatial cues made collaboration significantly more efficient by speeding up the process, lowering the total error rate (mismatches in memory task), and improving the overall score. The score of a task was measured by number of attempts (identification task) or the total deviation of the set cubes to their original positions (positioning task).

In order to measure the task completion time for each dyad in each task, we processed the data by detecting and excluding all time spans during the tasks where the game was paused. For the memory task we then only considered the time from the first to the last player interaction (cube opening) to delimit the actual runtime of the task. Before the first interaction, players did not know anything about the cubes' textures, that is why this point of time serves perfectly as a starting point. This does not apply to the second task - here, players were able to discuss their strategies and exchange their memories. Therefore, we chose the start of the game on the server as the starting point for the reconstruction task. The ending point was when both players pressed the "finish-button" and the game terminated.

To examine all possible indicators potentially being able to manifest influences of spatial cues, we extracted the average walking distance players in each room covered within each task. Therefore, the log data was sorted respectively by room and task. Then, paused parts were removed. After that, pairwise distances were calculated - from each position to the respective previous position. If they were smaller than a certain threshold they were summed up to the total distance. This threshold was necessary as the devices sometimes lost their position-matching and had drifting positions.

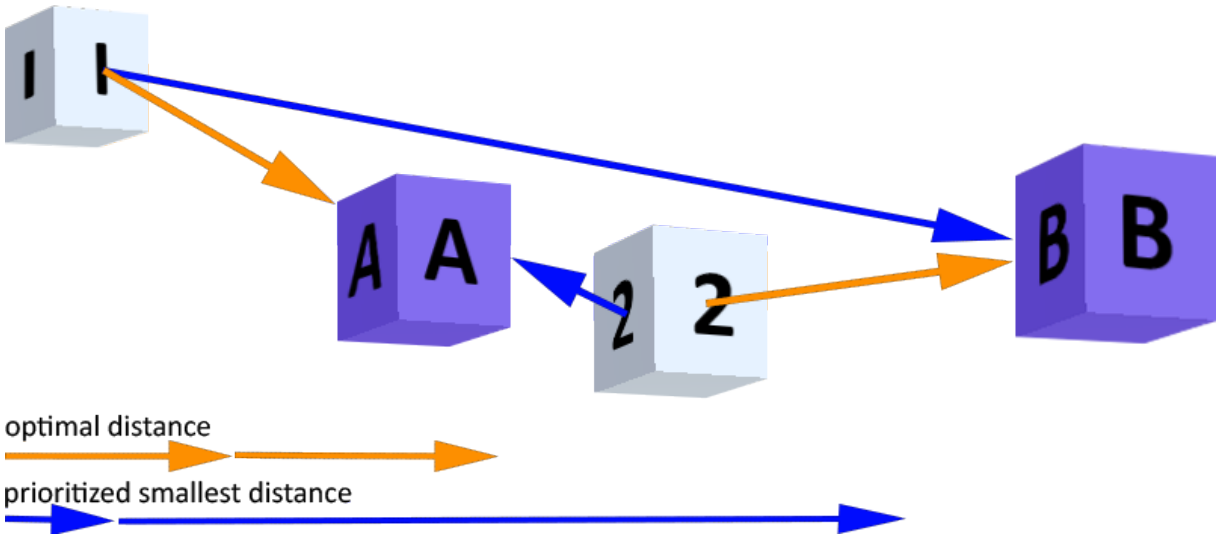


Figure 4.16: Reconstruction - deviation distance measurement. A and B are the original positions of a pair of cubes with the same texture. 1 and 2 are the respective cubes set by players during the reconstruction task. The groups' pairwise deviation can be calculated on two different ways: take optimal total distance (orange) or prioritize the smallest distance (blue).

In order to measure the groups' scores in the reconstruction task, a universal measurement is needed. The distance of the set cubes to the original cubes qualifies best for this purpose. As illustrated in Figure 4.16, there are two possible approaches to calculate the deviation of one pair of cubes. As it is not specified by the users which of the two cubes they actually intended to set, the most probable combination has to be chosen manually. E.g., in Fig. 4.16 cube *2* (set by player) could be assigned to both original cubes A and B with the same texture.
It is either possible to choose the smallest total distance (Figure 4.16 orange) or to prioritize the smallest distance (Figure 4.16 blue). In the second case (blue), first the smallest possible matching of cube and coordinate (Figure 4.16: $2+A$) is chosen. Then, the remaining pair of coordinate and cube is also matched (Figure 4.16: $1+B$). After one strategy is chosen, all distances of the respective task are calculated and summed up. For our evaluation we have chosen the optimal distance as measurement for all distances. The reason for this is that several groups did not manage to estimate the spatial depth properly. Mainly, this affected groups that used to stand firm at one end of the room during the whole memory task. Therefore, their complete reconstruction was shifted in a certain direction. If we had chosen the prioritized shortest distance function, most of their pairs would have been matched exactly contrary to the players' intention.

Finally, we also created density/heat maps indicating where in the room players mainly moved around. Therefore, the log data was merged for each room by task. Then, only X and Z coordinates (horizontal plane) were considered and processed in R to receive an useful visualization. The results are top-down views of the room showing, with different colors, where the players were primarily located.

4.3 Results

In this section, the results of our study will be presented. After that, they will be discussed and explained in section 4.4.

4.3.1 Work Load

Quantitative results of the TLX questionnaire[2] are displayed in Figure 4.17. For the memory and the reconstruction task, the average rated values with and without cues are lined up for each rating scale. In order to come to conclusions the differences between task load dimensions under various conditions have to be statistically significant. In the following, it is differentiated between within-subjects factors and between-subjects factors.

Within-Subjects Factor - Cue (With/Without)

Object identification In the memory task, pairwise comparisons between the condition with and without cues did not reveal any significant differences. The visualization of pairwise averages rated in the respective scales (see Figure 4.17 top) even purports the trend that spatial cues increase users' task load (higher scores with cues in all scales but in performance).

Object positioning In the reconstruction task, three of the six scales (performance, effort, frustration level) were normally distributed and therefore evaluated with the parametric t-test. The remaining three (physical demand, temporal demand, mental demand) showed up to be not normally distributed and thus were evaluated with the Wilcoxon signed-rank test. The tests revealed two significant findings.
During the reconstruction task, participants felt that it was more effort to complete the task when no cues were present (M_{cues}=56.72, SD_{cues}=19.29, M_{no_cues}=62.19, SD_{no_cues}=20.90, p=.047, $t(31)$=-2.071). In addition to that, they estimated that it took longer for them to

[2]The complete questionnaire is attached in the appendix (Appendix A.1 - TLX Questionnaire)

complete the task when no cues were present (M_{cues}=40.63, SD_{cues}=19.91, M_{no_cues}=46.88, SD_{no_cues}=22.50, p=.048, Z=-1.980). Even though no more statistically significant tendencies could be found, the lineup of average values for each scale during the reconstruction task (see Figure 4.17 bottom) presages a possible trend towards positive effects of virtual cues on mental demand, frustration, and the total score.

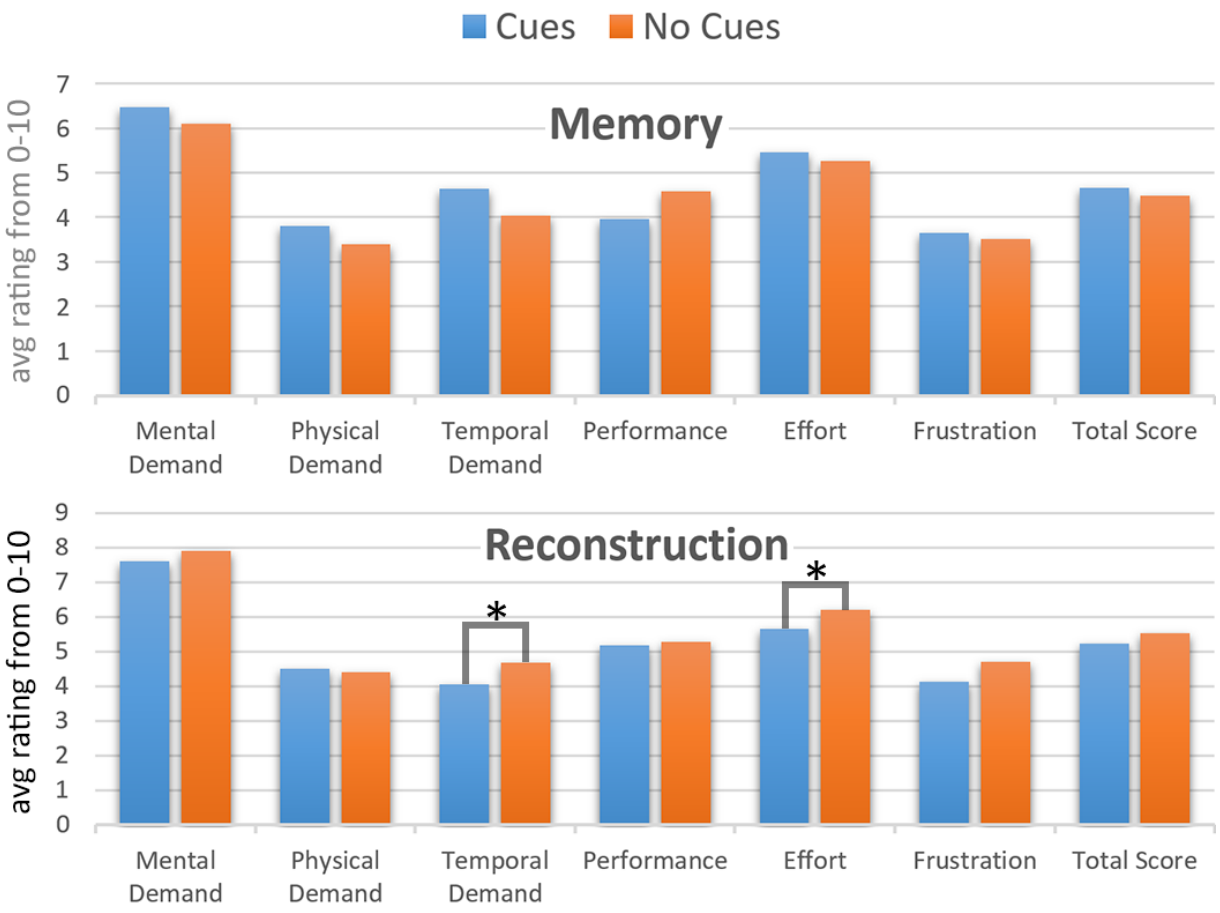


Figure 4.17: TLX quantitative distribution. Pairwise average ratings in the respective rating scales with and without the virtual cues.

Between-Subjects Factor - Room (907/923)

Object identification No significant differences between participants in the two rooms regarding participants' used spatial expressions could be found.

Object positioning In the reconstruction task when cues were present, the task completion in room 923 required a greater subjective temporal demand compared to the one in room 907 (M_{907}=35.00, SD_{907}=21.21, M_{923}=46.25, SD_{923}=17.37, p=.048, Z=-1.979; Wilcoxon signed-rank test). Except for this, no further significant differences between the two rooms could be detected.

4.3.2 Communication Behavior

To examine the communication behavior we categorized all occurring spatial expressions manually as described in section 4.2.4 (all categories used are listed in the appendix (B.1)). Figure 4.18 represents a rough overview of the quantitative distribution of used references. Each tuple of bars puts the respective number of references used without spatial cues in contrast with the amount of used expressions with virtual cues. Take note: in the following evaluation only differences in ratios of the variables are considered, not differences in the quantities themselves. E.g., Fig. 4.18: each bar shows the percentage of which the respective type of reference was used with regard to all other expressions with the same prerequisite.

General observations When cues were present we counted 18 more expressions in total. The number of occurrences varies even more, when only a smaller subset is considered (e.g., memory with cues: 862, memory without cues: 781). In the following, all values are taken percentage-wise to compare the differences properly in consideration of the different situations. In this overview the classes are still most detailed separated, thus it is differentiated between relative (blue) and absolute (red) references.
Obviously, deictic speech was used primarily to communicate. We counted 28% more deictic expressions when no virtual cues were present. Similar occurrences appear with regional references like *"further towards the window"* or *"higher"/"lower"*. When virtual cues were present the overall amount declined by 35%. Moreover, spatial cues were used in almost 21% of all expressions when they were present (memory: 25%; reconstruction: 17%).
To take a closer look at the groupings of all classes by object of reference, Figure 4.19 and 4.20 again contrasts results of different test situations with each other. Figure 4.22 allows a direct comparison of the tasks (memory/reconstruction) whilst taking the different test conditions (with/without cues) into account.

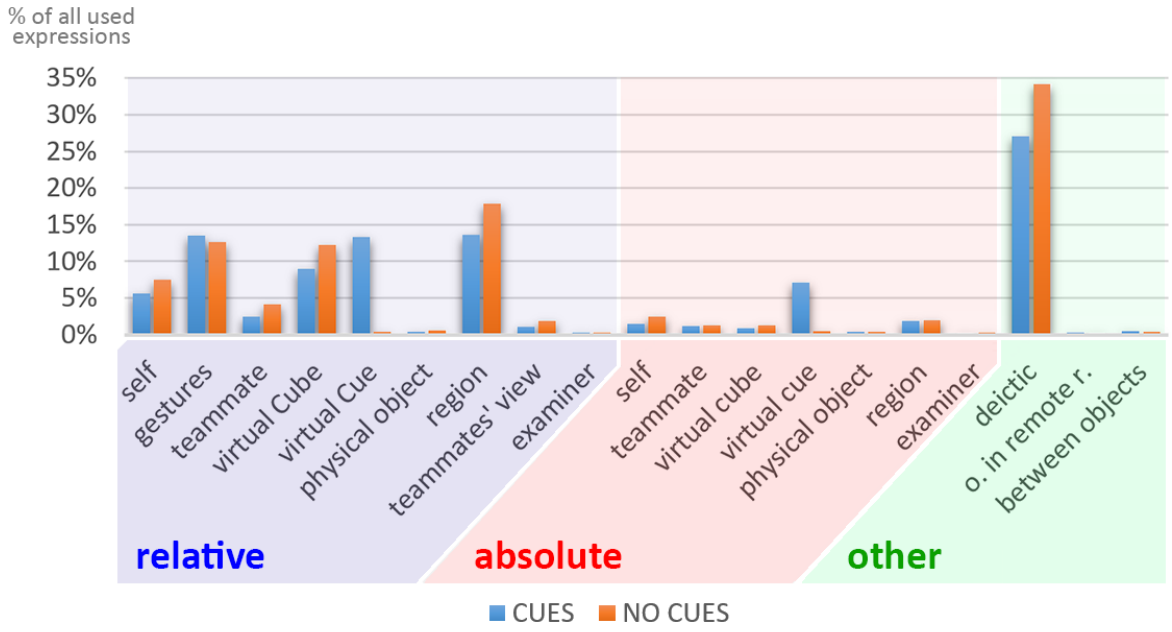


Figure 4.18: Overview: distribution of used spatial expressions (both tasks combined).

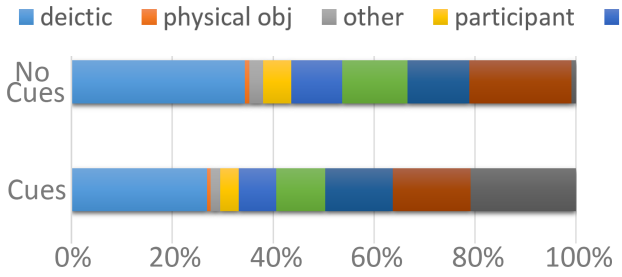


Figure 4.19: Distribution of references with and without spatial cues.

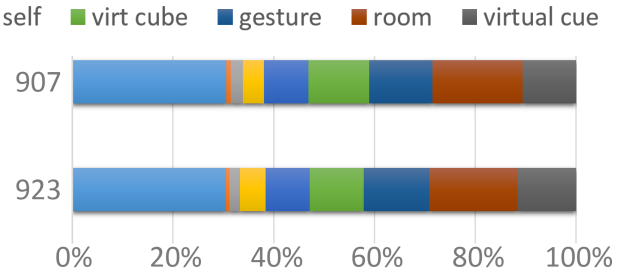


Figure 4.20: Distribution of references in the different rooms.

Considering Figure 4.19, an interesting phenomenon is that if one looks closely at the rightmost dark gray areas of "virtual cue" he will recognize that also under the premise "No Cues" there are references to virtual cues. Observation explains that this is actually due to the fact that several test persons referred to positions using virtual objects from the previous study condition in which virtual objects had been provided. Only half of the dyads did the two tasks with cues first and after that without, and the other half vice versa (Appendix A.4 - parameter settings). Therefore, there might have been even more references like that under different circumstances.

In Figure 4.20, the used references are grouped by room. As a reminder - room 907 was the messy room with many real visual cues. By contrast, room 923 was clean and plain, featuring only a couple of real spatial cues. The chart shows that the overall trend was the

same. Nevertheless, several differences can be noted. Participants in the plain room tended to use more deictic speech and references to virtual cues, whereas persons in 907 used the room and gestures more often for pointing out locations of interest.

Object identification In order to interpret those characteristics statistically correct, the Wilcoxon signed-rank test was applied on the data. During the object identification task, participants used significant more relative references to their teammates (M_{cues}=1.438, SD_{cues}=2.00, M_{no_cues}=2.500, SD_{no_cues}=1.93, p=.026, Z=-2.221) and to their teammates' view when no cues were present (M_{cues}=0.188, SD_{cues}=0.40, M_{no_cues}=1.375, SD_{no_cues}=2.00, p=.019, Z=-2.352).

Object positioning In the reconstruction task subjects used significantly more relative expressions to themselves (M_{cues}=2.31, SD_{cues}=2.02, M_{no_cues}=4.688, SD_{no_cues}=3.14, p= .003, Z=-2.947) and to previously set memory cubes (M_{cues}=6.625, SD_{cues}=4.015, M_{no_cues}= 10.188, SD_{no_cues}=7.96, p=.038, Z=-2.079) when no cues were present. Furthermore, they tended to use more absolute references to themselves - even when no cues were present (M_{cues}= 0.688, SD_{cues}=0.60, M_{no_cues}=1.625, SD_{no_cues}=2.09, p=.041, Z=-2.043).

For both tasks, participants referenced the visual cues significantly more when they were present (object identification: p(relative)=.001, p(absolute)=.001; object positioning: p(relative)=.000, p(absolute)=.001).

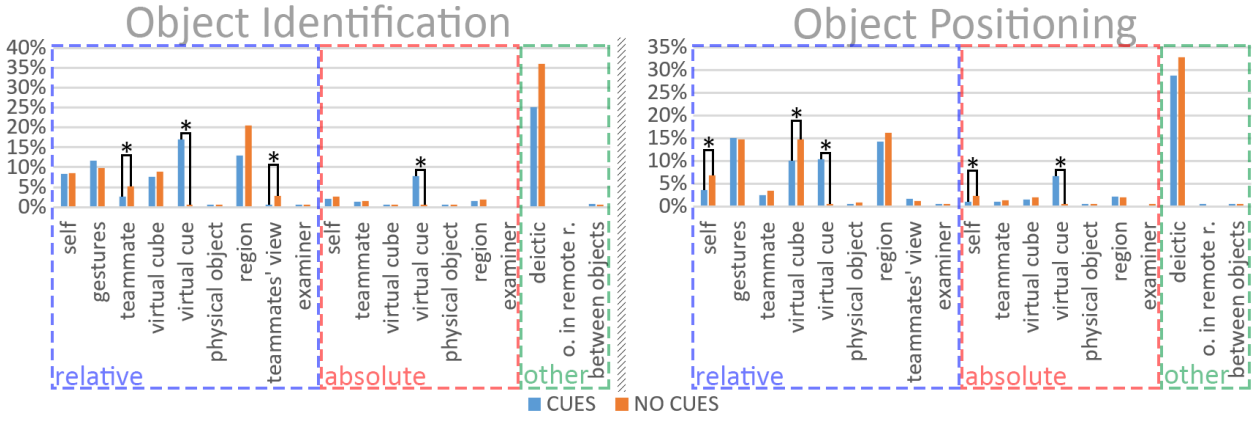


Figure 4.21: Communication behavior. Used expressions by task.

Figure 4.22 provides an overview of the distribution of references when grouped by object of reference. Using this grouping, the type of reference (relative or absolute) is neglected. Moreover, this diagram shows that similar trends in the results of the memory task compared to the ones of the reconstruction task were obtained.

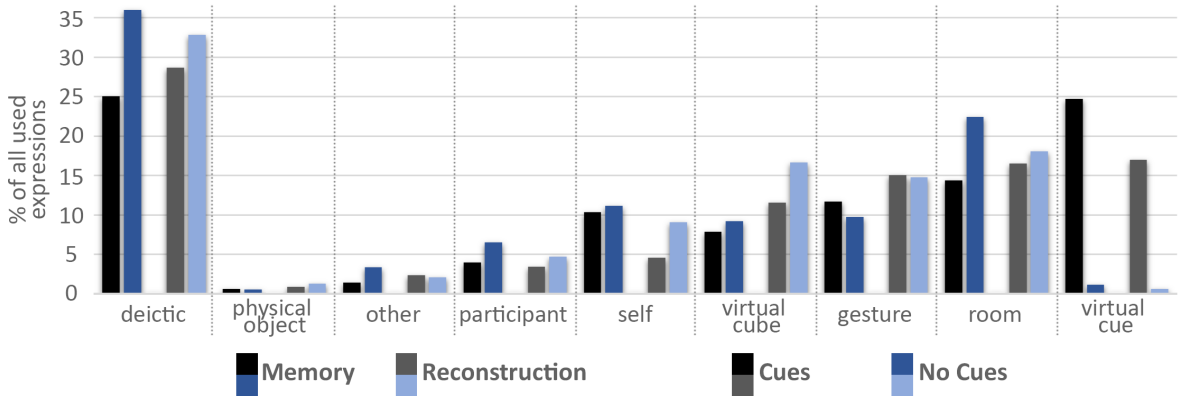


Figure 4.22: Distribution of grouped, used spatial expressions by tasks.

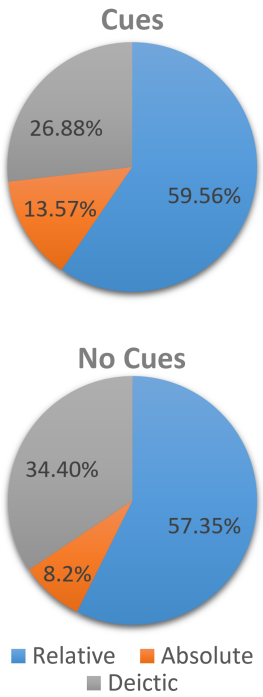


Figure 4.23: Proportions of relative and absolute references with and without spatial cues.

Relative & absolute references Figure 4.23 illustrates the distribution of used relative and absolute references with and without cues. It shows that when spatial cues were present about 27% of all used spatial expressions were deictic and about 14% were absolute. This relationship changes significantly as soon as the cues are removed. The total amount of used deictic speech rises by about 28% to over a third of all used references. At the same time, the total amount of absolute references declines from 40% to 8%. Accordingly, virtual cues have no major influence on the total amount of relative references - but doubtless on the objects that are involved in those (Fig. 4.19 makes that clear).

References to real objects One more interesting remark is that most participants intuitively assumed that both see the same virtual objects as oneself. Only two groups first matched what they see to check if everybody sees the objects at the same places. However, others even thought that their rooms look exactly alike and referenced real objects of their environment. Sometimes that caused problems, as several things existed in both environments, but were at completely different locations (door, notebook, desk...).

Memorization strategies By no later than the second tuple of tasks (memory + reconstruction) most groups decided to proceed on using some strategy to improve their score. The most popular one was to supervise the whole area from the same point of view. I.e., both players stand on the same side of the room and open the cubes alternately each after another. Another one was to divide the room into two halves, each player being responsible for memorizing the cubes on one side. Others made logic concatenations between cubes alone or between cubes and virtual landmarks. E.g., "schoolchildren take the bus; schoolchildren paint" is a logic concatenation used to link the memory cubes with the textures "Bus" and "Pen". So when they would remember one of them, they could tell that the other one was close by. Different approaches were, for instance, to tell a story using all the pictures in their respective order where they are located at (e.g., from left to right; top to bottom), or to split up memory pairs so each team member would have to memorize one of the cubes.

4.3.3 Presence - Extended Temple Presence Inventory

The quantitative analysis of the extended TPI[3] data delivers a quick overview. Averages of all scales but two are greater when cues are present. The first inverted scale was of the question: "How often did you want to or did you make eye-contact with someone you saw/heard?" (1 Never - 7 Always). The second one was a two dimensional rating scale - also from 1 to 7: "Unresponsive" at 1 to "Responsive" at 7. An overview of all results is provided on page 68 in Fig. 4.24 (for more details see the diagram of all averages during the memory task: Appendix B.1 - TPI: average values - object identification task).

Object identification To test the differences' significance we used the Wilcoxon signed-rank test. First of all, both of the above mentioned differences did not turn out to be statistically significant and therefore receive no further consideration. During the object identification task, two significant characteris- tics could be identified.

The first one was that participants ranked their ability to communicate spatial infor- mation with their partners higher when cues were present (newly introduced question - not part of the original TPI questionnaire):

How were the possibilities to communicate spatial information with your partner? (very bad 1 - 7 very good)
(M_{cues}=5.844, SD_{cues}=0.92, M_{no_cues}=5.156, SD_{no_cues}=1.63, p=.025, Z=-2.234).

[3]The extended TPI questionnaire is also attached in the appendix (Appendix A.2 - TPI Questionnaire)

Second, when cues were present they rather felt as if they were transported - together with their teammate - to another location. This implies that with spatial cues participants felt more present in the virtual part of the MR than without cues:

How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place? (not at all 1 - 7 very much)
(M_{cues}=4.000, SD_{cues}=1.70, M_{no_cues}=3.406, SD_{no_cues}=1.76, p=.035, Z=-2.104).

Object positioning These two questions yielded significant differences in the object positioning task as well.

How were the possibilities to communicate spatial information with your partner? (very bad 1 - 7 very good)
(M_{cues}=5.656, SD_{cues}=1.00, M_{no_cues}=4.844, SD_{no_cues}=1.55, p=.003, Z=-2.943).

How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place? (not at all 1 - 7 very much)
(M_{cues}=4.188, SD_{cues}=1.94, M_{no_cues}=3.438, SD_{no_cues}=1.23, p=.021, Z=-2.308).

During the reconstruction task five further significant differences emerged. First, when cues were present participants felt rather to be in the same room as their co-worker:

How much did it seem as if you and the people you saw/heard were together in the same place? (not at all 1 - 7 very much)
(M_{cues}=5.281, SD_{cues}=1.20, M_{no_cues}=4.125, SD_{no_cues}=1.83, p=.001, Z=-3.261).

Furthermore, when cues were present they had the feeling that they were more in control over interactions with their co-worker:

Seeing and hearing a person through a medium constitutes an interaction with him or her. How much control over the interaction with the person or people you saw/heard did you feel that you had? (none 1 - 7 very much)
(M_{cues}=5.000, SD_{cues}=0.92, M_{no_cues}=4.500, SD_{no_cues}=1.32, p=.023, Z=-2.282).

The last three were two dimensional rating scales, which all had the same instruction:

For each of the pairs of words below, please circle the number that best describes your evaluation of the media experience:

IMPERSONAL 1 - 7 PERSONAL (M_{cues}=5.063, SD_{cues}=1.52, M_{no_cues}=4.719, SD_{no_cues}=1.33, p=.012, Z=-2.524).
INSENSITVE 1 - 7 SENSITIVE (M_{cues}=4.938, SD_{cues}=1.00, M_{no_cues}=4.469, SD_{no_cues}=1.20, p=.037, Z=-2.083).
DEAD 1 - 7 LIVELY (M_{cues}=5.625, SD_{cues}=0.99, M_{no_cues}=5.09, SD_{no_cues}=1.21, p=.009, Z=-2.601).

Together, the evaluation of the three rating scales indicates that, when cues were present, participants' overall media experience was more personal, sensitive, and lively.

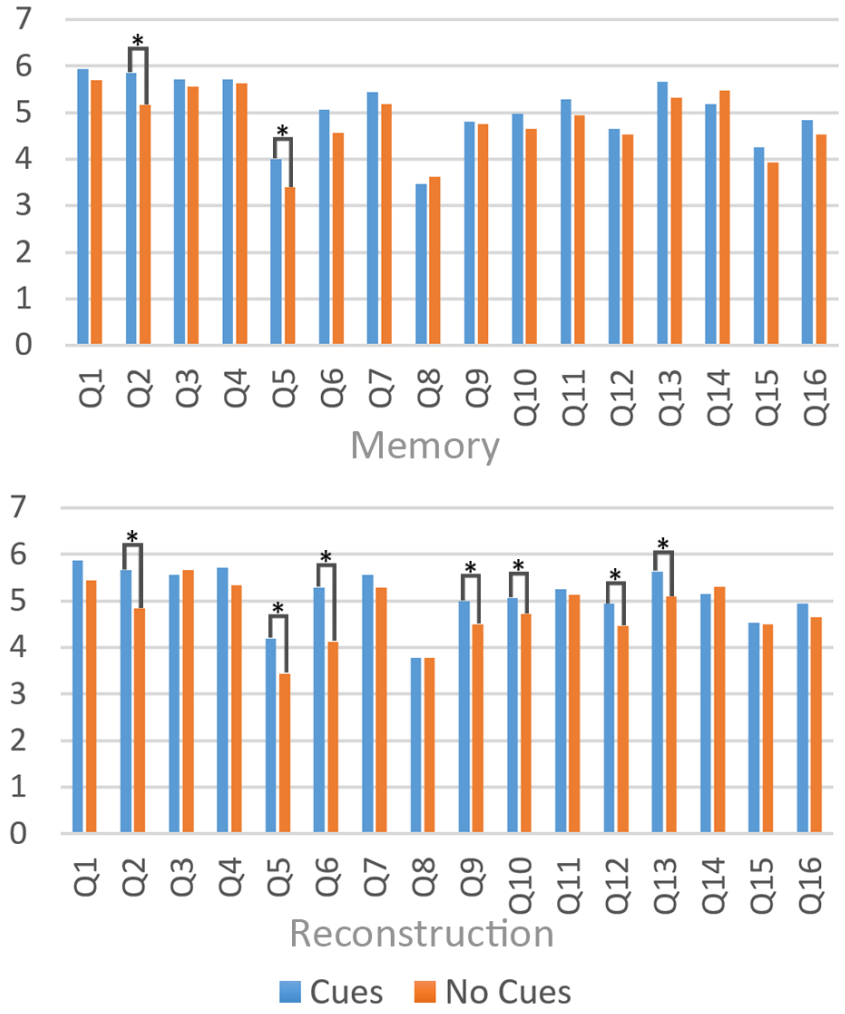


Figure 4.24: Extended TPI - results. Labels listed on page 69.

Labels for Figure 4.24 - questions of extended TPI:

Q1	*How were the possibilities to coordinate actions between you and your partner?*
Q2	*How were the possibilities to communicate spatial information with your partner?*
Q3	*How often did you have the sensation that people you saw/heard could also see/hear you?*
Q4	*To what extent did you feel you could interact with the person or people you saw/heard?*
Q5	*How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place?*
Q6	*How much did it seem as if you and the people you saw/heard were together in the same place?*
Q7	*How often did it feel as if someone you saw/heard in the environment was talking directly to you?*
Q8	*How often did you want to or did you make eye-contact with someone you saw/heard?*
Q9	*Seeing and hearing a person through a medium constitutes an interaction with him or her. How much control over the interaction with the person or people you saw/heard did you feel that you had?*
Q10-16	*For each of the pairs of words below, please circle the number that best describes your evaluation of the media experience.*
Q10	*IMPERSONAL - PERSONAL*
Q11	*UNSOCIABLE - SOCIABLE*
Q12	*INSENSITVE - SENSITIVE*
Q13	*DEAD - LIVELY*
Q14	*UNRESPONSIVE - RESPONSIVE*
Q15	*UNEMOTIONAL - EMOTIONAL*
Q16	*REMOTE - IMMEDIATE*

4.3.4 Interview

Preferred condition The script of the semi-structured interview is attached in the appendix (Appendix A.3 - Interview Structure). At the beginning, we asked several standard questions to receive some overall feedback on participants' opinions and tendencies. For instance, we wanted to know from participants which situation they preferred with regard to the display of cues. Without exception, everybody stated that the spatial cues were a great help and that they prefer to have them present during the tasks.

As an explanation for this, they named different reasons (several groups indeed named multiple). The most frequent cause - mentioned by 14 groups - was that the virtual cues enhanced their orientation. The additional cues served as universal pointers or fix points, that both team members could use to navigate each other. The words "common grounding" (in German: "gemeinsame Grundlage") occurred multiple times (n=5). Several participants mentioned that the enhancements did particularly apply because the cues stood out visually, as they were not rendered photo-realistically enough (n=3). In their opinion, that was the main reason for their team to relate to them so frequently, and not primarily because of the common grounding. Furthermore, the cues supposedly provided an opportunity to conveniently reference even more cubes by their clear position near a virtual object (i.e., "above the bookshelf"). Another reason why the cues enhanced the players' orientation was that they eased the estimation of differences in distance and height. As several groups described, by using those virtual cues with their distinct size and position as some kind of measure, they rather were able to estimate the distance and height of memory cubes than without (n=9).

Another explanation of several participants, why they preferred the tasks with the cues present, was that they enhanced their (spatial) memory capabilities (n=12). According to their statements, creating logical connections between the textures on the memory cubes and close by virtual cues improved their capabilities to memorize and remember the cubes' textures and positions. Like that, it was easier for them to find matching pairs in the memory task and to position them afterwards more accurately in the reconstruction task. To give an example, one participant kept the concatenation "the plane flies over the forest" in mind. In this case, the memory cube with the plane symbol was positioned over one of the plants.

Importance of cues by task With regard to the next question - if virtual cues were more important for one task than for the other or if they were equally important - the majority (n=27) claimed that the cues were much more important for the reconstruction task, but still were also a help for the memory task. About 22% stated that they found the cues to be equally important for both tasks. On the other hand, about 19% said that the cues had no influence on the memory task at all.

Real environment During the interview, each participant was prompted to choose a number between 1 and 10, indicating how important he or she thought the real environment was for him or her. The results are shown in Figure 4.25. The average value chosen was 5.38 (SD=2.60, M=5.5). The majority (about 60%) chose a value between 5 and 8. About 30% did not pay much attention to the real environment (1-3), whereas roughly 10% stated that the real environment was extremely important to them (9, 10). Approximately 90% of all participants mentioned that the real surroundings became much more important in the reconstruction task; the remaining 10% perceived it to be equally important in both tasks.

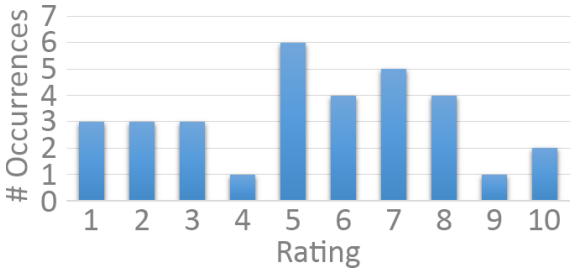


Figure 4.25: Interview evaluation. Participants' average rating on how important the real room was.

Reasons named by persons who rated 5 or more were, for instance, that the real environment was a central point of reference for orientation and that it was crucial as well for memorizing the positions of memory cubes. Besides, some mentioned at this point that they used real objects to create concatenations between these and memory cubes. Another point, which was relatively frequently mentioned, was that the walls of the room were absolutely fundamental in order to orientate properly. According to that, e.g., in the center of a huge sports hall, the game would have been much harder as distances and positions would be difficult to estimate without walls. Someone noted that the real environment was important, because it did not change at all and was a continuously existing static frame of reference, whereas the virtual room had supposedly a slight drift sometimes and changed its contents (virtual cues).

Participants with the opposing view - i.e., that the real environment was not that important - also had different arguments for their opinion. Most stated that the reason for them to discard and ignore the real environment was, that it was no common grounding for both players. Thus, they focused completely on the virtual component and only considered the real environment subconsciously. Some mentioned that the real room did not make a difference at all and one participant even suggested that they could have been "located in a completely dark room" as well. Another person remarked that he felt that the real environment was rather distractive and would have been better off without it.

Correlation between virtual cues and importance of real environment On the following question, if the 'value' of the real room was somehow conditionally related to the virtual cues, over two thirds responded that, in their opinion, the real world was less important when virtual cues were present. The virtual landmarks functioned as "better and stronger references" and were common points of reference for both players. Some brought

up that real objects, like the touch table in 923, were replaced by virtual objects (in that case by the plant standing on the touch table). The remaining third, which found the real environment to be equally important under both conditions, argued that the virtual objects were blended very well in the real environment - so they were just a further help by posing additional points of reference.

Important objects The preferred physical objects of the participants in room 923 were the touch table (n=10), the cupboard (n=2), and the ceiling light (n=2). In 907 the favorites were: table (n=3), shelf (n=3), and study examiner (n=2; counted here as an "object", as he sat at the same place the whole time). Preferred virtual cues in both rooms were the plants (n=17), the bookshelf (n=8), and the chandelier ceiling lamp (n=5).
Reasons for this choice were, above all, their outstanding characteristics (unreal, non-realistic lighting), their capabilities to divide the room ("side of the door") and the distance to memory cubes (e.g., when a memory cube was close by, the virtual landmark gained in importance). Again, most participants found that all named objects were more relevant during the reconstruction task.

Suggested useful features As a final question, we asked the participants what features they missed and how the game could be improved. Many interesting thoughts and ideas were mentioned. The most frequent one was to embed a three dimensional grid into the environment. With that, one could estimate distances and remember positions more easily. Another suggestion was to add some kind of pointer to the toolbox of each player. Hereby, one could point on cubes and mark them without having to navigate the other's attention to that place by describing the position by words. Moreover, several groups mentioned that they wished to have had a greater field of view since it was very limited due to the small handheld screen. A further frequently heard suggestion was to implement a function, which enables players to see what the other one currently perceives - displaying it either, for a short moment, on the own tablet (see-through is temporary replaced) or continuously on an external device (e.g., wall mounted monitor).

4.3.5 Additional Findings

Score memory task (average attempts) Apart from our main information sources (videos, questionnaires, interview), we evaluated the log data from the tango tablets as well. As shown in Figure 4.26, participants needed fewer approaches to finish the memory task when cues were present (M_{cues}=18.375, SD_{cues}=2.58, Mdn_{cues}=18, M_{no_cues}=19.56, SD_{no_cues}=3.83, Mdn_{no_cues}=19). The required attempts ranged between 14 (yielded in the condition without cues) and 27 (again, without cues).

Task completion times The average completion time for the memory task was higher when virtual objects were displayed (M_{cues}=6min 31s, SD_{cues}=135s, Mdn_{cues}=6min 19s; M_{no_cues}=5min 51s, SD_{no_cues}=126s, Mdn_{no_cues}=5min 24s).
For the object positioning task, participants needed in the condition with cues on average less time (M_{cues}=7min 41s, SD_{cues}=154s, Mdn_{cues}=6min 34s; M_{no_cues}=7min 55s, SD_{no_cues}= 135s, Mdn_{no_cues}=7min 34s).

Score positioning task (total deviation from original constellation) The average position-deviation[4] in the reconstruction task was M_{cues}=26.572m (SD_{cues}=3.65m, Mdn_{cues}= 26.43m) with cues, and M_{no_cues}=25.403m (SD_{no_cues}=5.07m, Mdn_{no_cues}=24.98m) without cues. The best achieved outcome - i.e., the attempt closest to the original setting - was obtained while no cues were present (15m). The same holds for the worst attempt (worst while no cues present: 34.05m; worst while cues present: 33.6m).

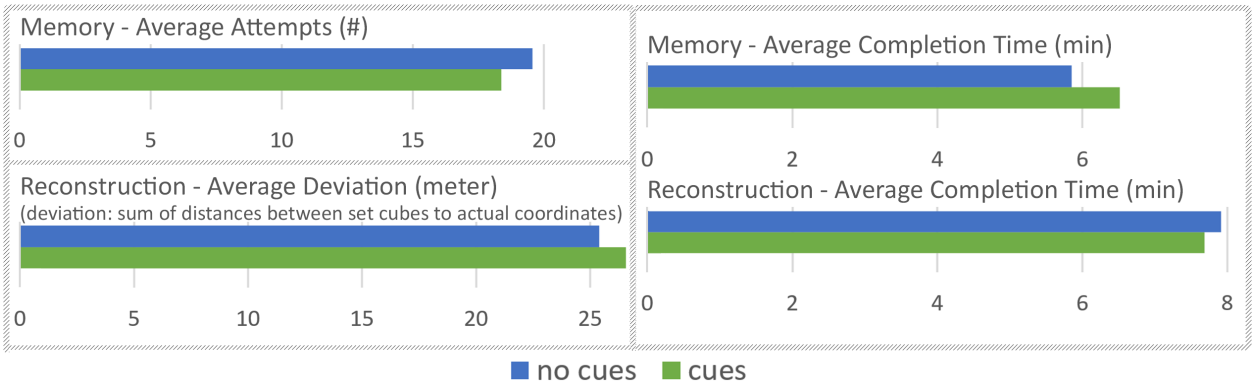


Figure 4.26: Descriptive results of the log file evaluation. "Scores" in memory task (number of attempts; top left) and reconstruction task (total distance; bottom left). Right: average completion time depicted separately for each task.

[4]understood as the distance between is-positions and should-positions of the set cubes. For a detailed explanation see page 59

Further statistical tests, which were carried out, indicate no significant differences regarding the above mentioned variables (attempts, deviation, completion times) due to changes in the factor cue (with/without) or room (907/923).

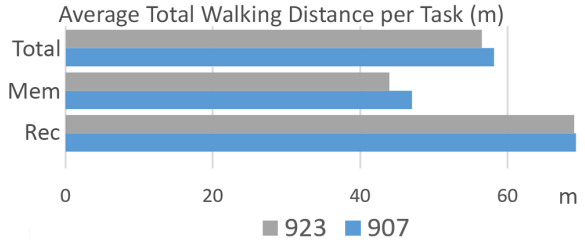


Figure 4.27: Average walking distance in meters for each room and task.

Players' walking distance As shown in Figure 4.27, the average walking distance was quite similar in both rooms, only slightly differing in the memory task. With regard to the factor cue (with/without), the average covered distance did not vary significantly as well.

Heat maps In order to be able to relate, where in the room players mainly moved around, heat maps were created - for each room and task separately. They reveal that subjects in the "plain" room (923) apparently moved around more and that they did not remain at a certain position - like most participants in the messy room (907) did (Fig. 4.28). Moreover, with respect to the different tasks, they suggest that players moved around more in the reconstruction task compared to the memory task. All six visualizations (each room: memory, reconstruction, and a combination of both tasks) are attached in the appendix (Appendix B.1 - Heat map: Player Positions).

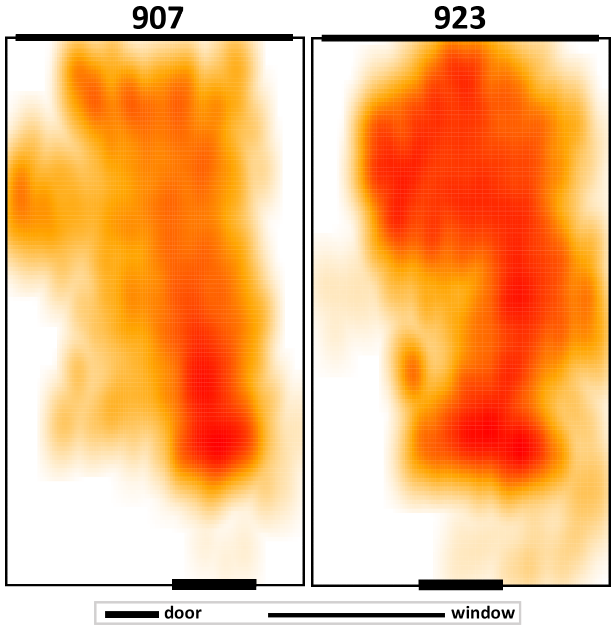


Figure 4.28: Heat maps of room 907 (left) and 923 (right). Top down view on rooms. Colors indicate where players moved mainly (red) or rarely (white).

4.4 Discussion

The evaluation of our four main dependent variables and several additional factors did not deliver an unambiguous and clear result. Rather, it provides a broad margin to discuss and interpret possible trends and implications. In this section, several interesting results will be picked out and discussed with regard to their validity in order to draw conclusions in terms of collaboration performance.

Communication Behavior

Overall, the video analysis indicated several significant differences between the various examined study conditions. During the memory task participants referenced themselves (with relative references) more often when no cues where present (e.g., "in front of me"). Despite there were only few self-references in total, this could imply that the collaborators had a smaller common grounding when no cues were provided and therefore had to fall back on self-references due to the lack of alternatives. The same holds for the second significant characteristic in the memory task: participants used more relative references ensuing from their teammate's position (e.g., "to your left") when no cues were present.
What further supports the assumption mentioned above is the observation that during the object positioning task virtual cubes gained in importance as soon as spatial cues were absent. Virtual cues were, when available, a central point of reference. In the memory task subjects used them in 25% of the cases. When no cues were present this large amount was compensated partially by the more frequent use of self-references and references to the teammate's view instead of using the real surrounding and its spatial features. The reason for this is that the shared common grounding is relatively small and restricted to the participants themselves (abstract models), the game objects (reconstruction cube, memory cubes) and, if available, spatial cues. Thus, by taking away the cues, a huge fraction of the common grounding is removed as well - the players have to resort to the remaining two classes (cubes, player models) or deictic speech.

Besides that, several trends could be observed which did not reach significance in the statistical tests. As our sample size was rather small, only very large effects could be detected statistically. In the following, findings which turned out to be not significant in a statistical sense, are discussed as well since they may be practically relevant (and detectable with larger sample sizes).

There was a scant increase in individual, inconclusive and, with regard to the real environment, ambiguous references (e.g., physical objects) in case cues were not displayed. Nevertheless, most changes appeared in other categories which were part of the common grounding (virtual cues, player model) or universal expressions (deictic).

During the object identification task, when no cues were present, participants tended to use deictic speech more often (Fig. 4.23). A possible explanation for this is the lack of a distinctive common grounding. As Clark and Brennan (1991) explained in their paper, a common grounding is fundamental for collaborative communication. In addition they suggested that people tend to use minimal effort in order to successfully convey their expressions. Therefore, Clark and Brennan distinguish between eight different "conventional" mediums - each needing its own kind of common grounding. For instance, in a co-presence scenario (medium 1) two share the same physical environment and can refer to everything around them, whereas during a telephone conversation (medium 2 - audibility) only timing and intonation are conceivable indirect communication tools. Everything, like pointing out a certain position or taking reference to a certain object, has to be explained with words which are independent of the setting one is situated in. Taking unexplained references to external appearances and making gestures or glances has no effect without verbal description. Since their article dates back to the early 90s, Clark and Brennan did not take the blending of different mediums into consideration. In our case, we created a 'semi-co-presence' scenario: both participants can perceive each other's presence and can talk to each other, but only share a part of their surroundings (the virtual component of the MR). Nevertheless, a part of their conclusion still fits and can be transferred to our situation. In our application the minimal effort, to point out a certain situation without having virtual cues, consists in using deictic expressions in combination with gestures (using the abstract player model). During the experiment often just the own position was used (*"here!"*) to point out a specific cube (the closest to a player's model). At other times subjects used the model's frustum (field of view) to specifically point on one cube (*"this one!"*). Therefore, spatial cues might extend and enhance the common grounding enough to have an impact on collaboration.

The evaluation of the interview provides similar conclusions. Most participants (n=21) stated that the reason why they preferred the situation with spatial cues was that hereby they had common points of reference and therefore a better orientation and enhanced navigation abilities. Furthermore, several subjects even used the term "common grounding" and mentioned that the cues were a great enhancement of the common communication basis. Apparently, they found it difficult to compensate the missing cues and, in some cases, actually utilized their old positions as a reference, when the cues no longer were displayed (video analysis).

A comparison of all employed expressions with regard to the different characteristics of the rooms reveals some trends, too. By way of example, the percentage of overall used references to virtual cues was higher in the "plain" room 923. This could be due to the fact that the room did not provide as many physical references as the other one and therefore hampered the usage of expressions which refer to the room itself. Further, it supports the suggestion that the lack of physical features rather leads to the blinding out of the real environment, whereas feature-rich surroundings enhance the continuous self-perception in the physical world.

Task Load

In contrast to Müller et al. (2015) we could not detect any impact of spatial landmarks on users' task load during the object identification task. Likewise, the different characteristics of the rooms did not seem to influence the task load during the object identification task significantly. The reason for this could be that the overall performance, demanded from participants, was too facile (floor effect). Possibly, they felt relatively unchallenged during this task compared to the object positioning task (on average they rated each scale 0.59 points higher after the reconstruction task in comparison to the memory task). Another possible limitation poses the small set of 16 samples.
As stated in section 4.3 the evaluation of the NASA TLX questionnaire did reveal several significant effects of cues or rooms on users' task load during the object positioning task. Time and effort were perceived as 'more demanding' when no cues were present. Furthermore, participants tended to perceive the reconstruction task with cues present as being more time demanding in the 'plain' room than their partners in the 'messy' room. Maybe this is due to the smaller frame of reference in the empty room. Whereas the feature-rich room offered many real points of reference for memorization of the cubes' positions, the 'plain' room did not. Similar results may be obtained for the object identification task using a more challenging task and employing a bigger sample.

The NASA TLX exclusively measured the subjectively perceived task load. Other measures, like the users' completion times per task or their overall performance (operationalized through the number of attempts or the total deviation distance), can additionally be considered as representations of their actual work load. Therefore, we computed the average amount of attempts needed to complete the object identification task. When no cues were present participants needed more (+1.2) attempts on average. If, in future research, this trend proves to be significant, one could assume that spatial cues are capable of improving users' effectiveness, which again might also lower the task load.
During the memory task the average completion time was longer when cues were present. On the contrary, in the reconstruction task groups were faster when cues were provided. This might result from the circumstance that during the memory task users rather relied on their remembrance when cues were present instead of just going with (probably time-saving) 'trial and error' tactics. As the previously mentioned comparison of the mean number of attempts between the respective conditions revealed, groups were more successful when cues were furnished. But, for this achievement, more time was required. For the simple reason that spatial cues provided better (i.e., more memorable) points of reference participants could more likely remember the respective textures and positions of the cubes. This took longer as it requires time to exchange one's spatial memory with the partner before opening the next cube. In contrast, when no cues were present, subjects were inclined to open cubes randomly, as it was harder for them to remember related cubes. As several observations deducted from the video analysis showed and statements in the interview underlined, participants put more

effort in memorizing the memory cubes when cues were present. Supposedly, with cues it was easier for them to build mnemonics, since it was the only condition in which they had continuously a static common grounding during both tasks (player models move, memory cubes disappear - only spatial cues stay at their positions for both tasks). Thus, it seems likely that the additional virtual objects improved the participants' spatial memorability of other objects in the MR. On average groups finished the reconstruction task faster when cues where present. Even though their mean overall performance was worse (although not significantly), they must have been surer about the positions of the memory cubes in the previous game as they placed the cubes in less time. At least, it would be a possible explanation for this appearance. An alternative explanation would be that the players were increasingly distracted by the cues. That may be true, but this approach cannot explain the development in the second task in which participants required less time when provided with cues.

Moreover, during the interview participants were asked which scenario they preferred. All without exception stated that it was easier for them to complete the tasks with cues. One can conclude that their sense, that the presence of spatial cues eased the task, is equatable with a lowering of their retrospective subjective work load (even though their answers in the TLX questionnaire did not support this directly).

Considering heat maps of the participants' movements (Appendix B.1 - Heat map: Player Positions), the impression is given that in both rooms players moved more during the reconstruction task than during the memory task (this is also confirmed by the players' average walking distances). This is due to the fact that, in order to position the cues, they had to walk to the respective location. Therefore, this is a direct consequence of our task design. However, the differences between the two rooms, which emerged within one task, cannot be explained in this way, revealing the necessity to ascribe it to other causes.
Although the average of all measured walking distances did not differ significantly between the rooms, the heat maps show qualitative differences. Further research is necessary to determine whether this is actually due to distinct characteristics of the rooms. The data provided by the heat maps reflects that players in the 'plain' room 907 tended to prefer about six positions, standing there relatively still or changing between those. In contrast, participants in room 923 walked more consistently and steadily through the room. This might be caused by worse navigation capabilities in the 'plain' room 923. There are fewer physical landmarks which could help to estimate the height or distance of a virtual object. Therefore, players had to mill around more while their collaborators in the other room could just change their positions along the margin of the 'game area' to get different perspectives on the cubes (as the heat maps show, they had a tendency to stand close to the door or to the windows, permanently having an overview of most cubes - and just moving between those points). This suggestion is supported by several interview statements as well. Most participants who mentioned that they had difficulties to navigate and to estimate the height and distance of cubes properly were in room 923. Unfortunately, the sample was too small, and the

prototype not designed specifically enough to investigate this issue in detail. Therefore, one can only speculate with regard to potential connections between the rooms and the users' behavior. Further investigation is needed to verify whether a feature-rich room can improve collaboration in remote MREs due to the provision of points of reference which ease the estimation of distances.

Another aspect of the work load of participants, that can be considered, is the outcome of the modified TPI questionnaire. In both tasks, users perceived that the spatial cues eased their spatial communication efforts. For the requested tasks it was necessary to exchange much spatial information. Thus, one can count the subjective relief with regard to communication efforts as a simplification of the overall tasks and, therefore, as a decline of participants' task load.

Perceived Presence

The evaluation of the modified TPI questionnaire displayed several significant differences regarding users' subjectively perceived presence between the respective situations with and without cues. The common trend in both tasks was that cues increased participants' perception of presence (awareness of the virtual component of the MR and their partner within it). Most often, the average rating was higher when cues were present (this applies to both tasks). In the memory game, on two scales significant differences emerged between the conditions with and without cues. In the reconstruction task, the same holds true actually for seven scales.

First, during the memory task, participants stated that it was easier for them to communicate spatial information with their partners when cues were displayed (extended TPI questionnaire - question 2: *How were the possibilities to communicate spatial information with your partner?*). This might be due to the previously discussed fact that additional spatial cues extended the participants' common grounding as more shared visual points of reference existed. Therefore, the additional information enhanced their spatial communication abilities.

Second, when cues were present, participants perceived themselves more as if they were transported together to another location (extended TPI questionnaire - question 5: *How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place?*). I.e., more often than before they had the feeling that both (themselves and their collaborators) were at a third location. In this case, this third location is the virtual component of the MR. I.e., the cues provide a common grounding, which is not the physical environment of one of the collaborators. The participants' overall statements suggest the assumption that spatial landmarks improved the participants' sense of awareness of the shared component of the MRE. Being aware of the common surrounding of the shared

work environment is a central factor for good collaboration as Clark and Brennan (1991) concluded. According to them, self-awareness in the common environment is crucial for communication. The more developed the shared grounding is (and the more participants feel to be a part of it), the easier it is to communicate (less misunderstandings and unclear expressions).

Both indicators of a clear trend towards the enhancement of one's perception of presence owing to cues were also significant in the object positioning task. Additionally, for this task, five more characteristics could be shown. Participants rather felt as if they were in the same room with their partner when cues were displayed (TPI question 6: *How much did it seem as if you and the people you saw/heard were together in the same place?*). Compared to the above mentioned question (TPI question 5: *How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place?*), this item aims more at how a participant perceives the other player - not the third room. The outcome indicates that the cues tended to 'shrink' the distance between the collaborators. When cues were present they felt not so far apart anymore, but instead more co-located.
This suggestion is further supported by the result of the rating scale capturing how personal or impersonal the media experience felt. Participants assigned higher values (*impersonal*=1, *personal*=7) when cues were available, indicating that they perceived the communication relationship between themselves and their partner as more personal with cues. I.e., the communication felt more like a face-to-face conversation when the common grounding was larger.

The three remaining rating scales, which led to significant results in the reconstruction task, generated similar, although slightly lower, values in the memory task. Accordingly, virtual cues made the media experience more sensitive (TPI question 12), more lively (TPI question 13) and gave users more control over interactions in the collaborative environment (TPI question 9: *Seeing and hearing a person through a medium constitutes an interaction with him or her. How much control over the interaction with the person or people you saw/heard did you feel that you had?*). This again might be due to a larger common grounding together with a heightened self-awareness in the shared part of the MR. Entering the shared environment to a greater extent enhanced their orientation within it. Therefore, in this vein it was easier for them to keep track of the position of their collaborator.

User Experience

The interview provided the most insight into the overall user experience. By means of direct questions we could investigate whether and in which way the cues influenced participants. As additional sources, the questionnaire data and the log files were taken as a basis to draw conclusions. The interview clarified that participants preferred the condition with cues. Often cited reasons were that the cues served as shared points of reference (hereby extending the common conversational grounding) and that they increased the memory capabilities of

participants. Interestingly, most participants ($\sim 90\%$) stated that, in their opinion, the cues were considerably more important in the reconstruction task as compared with the memory task. This could be ascribed to the previously made suggestion that the memory task was too easy for participants in general (floor effect). For instance, by putting more memory cubes in the same area, participants would not be able to just roughly memorize their directions, but would have to remember their certain positions. Therefore again, they probably would rather have to rely on spatial cues to memorize those positions.

The expressed notions of participants can be supported by several measured criteria as well. As the significance tests of the TLX revealed, users perceived the completion of the object position task as taking more effort when no cues were present. Consequently, spatial cues helped to lower the overall effort by somehow facilitating the task. Besides, tendencies in the log data, like the decline of needed attempts during the memory task when cues were present, further confirm the participants' interview statements. Moreover, results from the TPI questionnaire match their statements as well. This shows up in the circumstance that the condition with cues was rated as more lively, personal, and sensitive.

4.5 Limitations

In order to exploit the investigated variables properly and to compare object identification with object positioning, the used tasks should be equally demanding. In our case, they were not. During the interview, participants stated that they found the reconstruction task to be more difficult. Several other measures underline this suggestion (e.g., the average completion times or the subjective task load). Different results in the memory and the reconstruction task might be caused by a floor effect in the object identification task. Possibly, finding matching pairs was too easy in comparison to the reconstruction task in which participants had to set the cubes themselves. It might be due to this, that participants did not require the additional aid provided by cues. Therefore, it might be reasonable to conduct a similar study with modified tasks (e.g., more memory cubes). It should be mentioned that on various occasions, for example the communication behavior (video analysis) or the task load (TLX), the obtained results were in line with established assumptions, if not necessarily significant. On this account, future research should examine the introduced issues using bigger sample sizes.

During the interview many subjects explained that multiple times they had difficulties to find remaining cubes. In addition to that, several people had problems to get along with the way the mixed reality was displayed on the tablets and to manage the transfer of the positions on the display into the real environment. Both problems could be solved, or at least reduced, by using head-mounted displays (e.g., Oculus Rift or similar). Like this, dyads could concentrate on the collaboration and solve the task without disturbances. In that case

the input medium (here: touch display) has to be replaced with a similar effective input method (additional controller, gesture recognition, etc.).
Another point, which should be addressed and is shown in the heat maps, is that participants had the tendency to stand close to one end of the room, either near the window or the door. This might be attributable to the reasonable desire to keep a position from where they can overview the whole scenario. By using a different kind of display, such as a head-mounted one, this might have been different as it probably would not have been so hard for participants to keep track of all the cubes - even while walking around.

In our study setup we used two rooms. Both were overlaid with virtual objects (with regard to the virtual room, the physical doors/windows were on the same side). Some people did not seem to realize that the other's room could have been completely different. They used the door or the window as if they were part of a common visual grounding. Presumably, if the physical room would have been completely different (for example with regard to size and orientation), there might have been a shift in communication behavior as people would have realized rapidly that spatial references with regard to the room are not effective. Since they used these expressions quite frequently, it would have been interesting to see what kind of references they would have taken to compensate.

Regarding the rooms, there was yet another relevant factor. We designed one room to be relatively plain and the other one to be rich of visual features (tables, chairs, tools, electronics...). During the interview many people in the "plain" room brought up the touch table to be the most important object in the room (including virtual objects as well). Taken as a whole the room had few features, but those which were available appear to be a lot more relevant than all of the objects in the other "feature-rich" room. To examine the influence of physical features on collaboration it might have been better if this touch table would have been removed completely.

4.6 Implications

During the different tasks of our study participants often used physical characteristics, like windows, walls, or the door, as reference points. In our case, the rooms had the same size and the virtual component was matched on the physical environments in both rooms in the same manner. Obviously, that was a large help for them as they used references to physical objects quite often (in 17% of all expressions during the whole experiment). Therefore, it might be helpful if one at least aligns basic measures (size, orientation) of the virtual overlay according to the rooms' physical characteristics. For example, it makes sense to choose the center of a real room as the center of the corresponding virtual room and to align the virtual environment to the room's orientation. In office environments, for instance, it

would probably be better suited to align the common (virtual) room along one wall, heading towards the windows than to align it with respect to the collaborators' desks.

Our results imply that physical features play a major role with regard to users' navigation and orientation abilities. One of our rooms contained many features whereas the other one was relatively plain. Participants of the plain room seem to have faded out and ignored the real room to a greater extent, making it harder for them to communicate without deictic speech. By contrast, participants in the feature-rich room used less deictic speech and more spatial expressions concerning the physical room. Physical objects also seem to support users' spatial memory capabilities. Interview statements reveal that many people used physical objects in the immediate vicinity in order to memorize the memory cubes.
As the touch table (in our study in room 923) was supposedly a too catchy and outstanding object, within the scope of further research a similar study with more extreme settings should be conducted. For instance, one room could be completely empty, forming an intense contrast to the other room which could be furnished with even more features than our feature-rich room. All used additional objects should be even more outstanding and eye-catching than the ones we inserted. Especially suitable would be objects that usually do not belong to this environment (e.g., a small statue of a purple elephant).

As several different variables indicate, the common grounding is enlarged and enhanced once the MRE's virtual shared component is extended by means of visual cues. Main indicators for this are differences in the perceived presence of the virtual room and a strengthened self-awareness within it when cues are present (TPI). Moreover, in the interview all participants stated that the cues provided useful support. This leads one to believe that spatial cues have a positive impact on collaboration. No diametrically opposite observations were made. Everything considered, one can conclude that embedding virtual cues in remote collaborative MR environments could serve as an enhancement of the work environment.

4.7 Future Work

The video analysis revealed that participants attempted to direct their teammates' attention to a specific memory cube by gestures in ways that could be referred to as unhandy and laborious. In order to better convey gestures and gazes the player's model could be replaced by either a more detailed model including the face and hands of its player or even by the whole body as a hologram-like masking. A potential approach realizing the last mentioned possibility has been developed by Prince et al. (2002) (see chapter 2). It seems obvious, but nonetheless has to be proven, that this can improve communication possibilities and hence the overall collaboration.

Furthermore, instead of using a hand-held device like the tango tablets, the same scenario could be played using head-mounted displays. By this means, users would not only get an

even better feeling for the mixed reality, but would also be able to communicate more naturally with the hologram-like representation of their teammate. An encouraging study comes from Kiyokawa et al. (2002) (see chapter 2). They came to the conclusion that head-mounted displays with optical see through are the most efficient mediums with regard to collaboration in MREs.

A repeatedly expressed opinion during the interview was that the synthetic landmarks were important because they were eye-catching (due to their outstanding appearance), not because they functioned as a common grounding. It would be interesting to see the differences in collaboration results using different sets of cues. Very photo-realistic cues, which might accidentally be taken for real objects, on the one hand, and sets of very abstract and outstanding items on the other hand (e.g., a red flying elephant). Results of this research could improve the selection of cues in future real-life practical use cases.

As already mentioned in the previous section (4.6), the nature of the rooms opens up possibilities for further research. First, the influence of the alignment of the virtual rooms with regard to the physical rooms could be investigated. One room could match the virtual room's borders exactly, while the other room could be aligned quite differently. For example, whereas the first room matches the x-axis with one wall and is headed towards the windows, the second room could be divided by the x-axis having its window sideways. Alternatively, it could be located in a hall or outside providing no physical borders. Second, the influence of the amount and the characteristics of features in the rooms could be investigated. More extreme scenarios than ours would be necessary in that case. Think of a scenario with two identical rooms from which one is stuffed with spatial landmarks while the other one is completely empty, being nothing but a white box from the inside.

Some of the numerous suggestions for improvement of collaboration in our tasks could be implemented and examined. Most improvements could probably be transferred to more common collaboration tasks. One particularly promising proposal, which was in fact cited by several attendants, is to extend users' abilities through some kind of marking tool. Thereby, they would be able to mark or point on objects and positions. There are many different possible approaches to find a solution for this. One is to replace the abstract model of the collaborators by their actual image, which has to be scanned and transmitted in real time. As shown by Prince et al. (2002), the fundamental technical realization is possible even nowadays. With this technique, participants could easily point at things by means of gestures. Additionally, in order to facilitate marking as well, gesture recognition, or specific touch interactions could be used to insert markers which have a fixed location in the virtual environment. Yet another mentioned option was the insertion of a 3D grid in the virtual overlay, dividing the whole room into invisible cubes (grid = edges of cubes are visible). Supposedly, this would help users to navigate in the room, being able to better estimate heights and distances. Of course, for all assumed "improvements" it has to be shown (with the help of further research) that they actually improve the overall collaboration.

5 Conclusion

Our basic concern was to ascertain if and to what extend additional virtual landmarks in MREs can have an impact on collaboration - especially in object identification tasks. Therefore, we built a study prototype which was dedicated to explore this issue.

The precursory study by Müller et al. (2015) already suggests that spatial cues can have an influence on remote collaboration, as they play a crucial role in co-located MREs. Our study completes the results obtained by Müller et al. (2015), replicating them for remote MREs. The qualitative analysis of participants' communication behavior emphasized the importance of spatial cues also in *remote* MREs. The same holds for the evaluation of the interview statements. Besides, we investigated how participants perceived the presence of their teammate and their environment (above all, the shared component in the MR). The evaluation showed that additional virtual objects made a huge difference regarding participants' awareness of the shared virtual environment and their ability to adapt to it.

Certainly, further research is indispensable in order to gain more insight into which combination of characteristics used for the cues (such as position, amount, appearance, or size) is best suited. Therefore, research questions should be very detailed and focus on their specific research objective (e.g., is it better to use photo-realistic objects or rather very abstract ones?). Besides, future research could concentrate on examining other aspects which possibly play an important role, such as familiarization effects (as one might rather integrate virtual objects when one adapted to them).

To keep aloof from speculations, there is evidence that spatial cues can be useful in remote collaboration tasks (particularly with respect to communication behavior, user experience, and the perception of presence). Hence, these should be included in similar future real-world applications. To bring it back to the beginning: The first-mentioned example featured two architects working remotely together on the same 3D model of a building. Adding virtual cues to their MR could potentially assist them in solving object identification tasks (e.g., *which of the trees in the model does he mean?*) and positioning tasks (e.g., *whereto does he want me to relocate it?*) more efficiently. The thereby extended common grounding might improve the overall collaboration (one could already consider it as an improvement in case participants subjectively feel better with cues).

On balance, the results of the study indicate that spatial cues can have positive effects on remote collaboration in MREs. This assumption is supported by the implementation and evaluation of an exemplary collaboration scenario which combined two commonly acquired tasks, namely an object identification and an object positioning task. Several observations

indicate, or at least suggest, that spatial cues can enhance a teams' communication, ease task solving processes, and improve users' overall collaboration experience. However, the extent to which visual cues can actually enhance collaborators' communication abilities and the collaboration process itself is an interesting topic for future research and should therefore get further attention.

Bibliography

Bajura, M., Fuchs, H., and Ohbuchi, R. (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In *ACM SIGGRAPH Computer Graphics*, volume 26, pages 203–210. ACM.

Benford, S., Greenhalgh, C., Reynard, G., Brown, C., and Koleva, B. (1998). Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)*, 5(3):185–223.

Benko, H., Ishak, E. W., and Feiner, S. (2004). Collaborative mixed reality visualization of an archaeological excavation. In *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium*, pages 132–140. IEEE.

Billinghurst, M., Cheok, A., Prince, S., and Kato, H. (2002). Real world teleconferencing. *Computer Graphics and Applications, IEEE*, 22(6):11–13.

Billinghurst, M., Clark, A., and Lee, G. (2015). A survey of augmented reality. *Foundations and Trends in Human-Computer Interaction*, 8(2-3):73–272.

Billinghurst, M. and Kato, H. (1999). Collaborative mixed reality. In *Proc. Intl Symp. Mixed Reality*, pages 261–284.

Billinghurst, M. and Kato, H. (2002). Collaborative augmented reality. *Commun. ACM*, 45(7):64–70.

Billinghurst, M., Poupyrev, I., Kato, H., and May, R. (2000). Mixing realities in shared space: an augmented reality interface for collaborative computing. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference*, volume 3, pages 1641–1644.

Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C., and Steed, A. (2003). Lessons from the lighthouse: collaboration in a shared mixed reality system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 577–584. ACM.

Bubaš, G. (2001). Computer mediated communication theories and phenomena: Factors that influence collaboration over the internet. In *3rd CARNet users conference, Zagreb, Hungary*. Citeseer.

Cassell, J. and Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5):519–538.

Chalon, R. and David, B. T. (2004). Irvo: an architectural model for collaborative interaction in mixed reality environments. In *MIXER*.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149.

Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, 13(2):259–294.

Clark, H. H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1):1–39.

Fussell, S. R., Kraut, R. E., and Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 21–30. ACM.

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., and Kramer, A. D. (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309.

Gergle, D., Kraut, R. E., and Fussell, S. R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interaction*, 28(1):1–39.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2009). D4ar–a 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of information technology in construction*, 14:129–153.

Google Android Studio. Android Studio. Retrieved February 18, 2016 from http://developer.android.com/sdk/index.html/.

Google Glass Developer. Google Glass. Retrieved February 6, 2016 from https://developers.google.com/glass.

Google Project Tango. Project Tango. Retrieved January 25, 2016 from https://www.google.com/atap/project-tango/.

Google Tango Explorer. Tango Explorer. Retrieved January 25, 2016 from https://developers.google.com/project-tango/tools/explorer/.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183.

Hwang, A. D. and Peli, E. (2014). An augmented-reality edge enhancement application for google glass. *Optometry and vision science: official publication of the American Academy of Optometry*, 91(8):1021–1030.

IBM Corp (2015). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Kato, H., Billinghurst, M., Weghorst, S., and Furness, T. (1999). A mixed reality 3d conferencing application. *Human Interface Technology Laboratory*.

Kim, S., Lee, G., Sakata, N., and Billinghurst, M. (2014). Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *Mixed and Augmented Reality (ISMAR), International Symposium*, pages 83–92. IEEE.

Kiyokawa, K., Billinghurst, M., Hayes, S. E., Gupta, A., Sannohe, Y., and Kato, H. (2002). Communication behaviors of co-located users in collaborative ar interfaces. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium*, pages 139–148. IEEE.

Kritzenberger, H., Winkler, T., and Herczeg, M. (2002). Mixed reality environments as collaborative and constructive learning spaces for elementary school children.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*, volume 5.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive psychology*, 28(2):103–174.

Lombard, M., Ditton, T., and Weinstein, L. (2009). Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*, pages 1–15.

Microsoft HoloLens. HoloLens. Retrieved February 24, 2016 from https://www.microsoft.com/microsoft-hololens/en-us/experience.

Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329.

Müller, J., Rädle, R., and Reiterer, H. (2015). Virtual objects as spatial cues in collaborative mixed reality environments: How they shape communication behavior and user task load. In *Proceedings of the 34th annual ACM conference on Human factors in computing systems (CHI16)*, volume 38, pages 1845–1850. ACM.

Oculus VR. Oculus Rift - head mounted VR-display. Retrieved February 2, 2016 from https://www.oculus.com/.

Ohshima, T., Satoh, K., Yamamoto, H., and Tamura, H. (1998). Ar2hockey: a case study of collaborative augmented reality. In *Virtual Reality Annual International Symposium*, pages 268–275. IEEE.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics*, 30(1):20–28.

Prince, S., Cheok, A. D., Farbiz, F., Williamson, T., Johnson, N., Billinghurst, M., and Kato, H. (2002). 3d live: Real time captured content for mixed reality. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium*, pages 7–317. IEEE.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, P. and Tuddenham, P. (2007). Distributed tabletops: Supporting remote and mixed-presence tabletop collaboration. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop*, pages 19–26. IEEE.

Rogers, Y., Scaife, M., Gabrielli, S., Smith, H., and Harris, E. (2002). A conceptual framework for mixed reality environments: designing novel learning activities for young children. *Presence: Teleoperators and Virtual Environments*, 11(6):677–686.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., and Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201.

TeamSpeak Systems. TeamSpeak 3. Retrieved February 18, 2016 from https://www.teamspeak.com/teamspeak3/.

Unity. Unity - Gaming Engine. Retrieved January 30, 2016 from https://unity3d.com/.

Unity Asset Store. Unity 3D Asset-Store. Retrieved January 31, 2016 from https://www.assetstore.unity3d.com/.

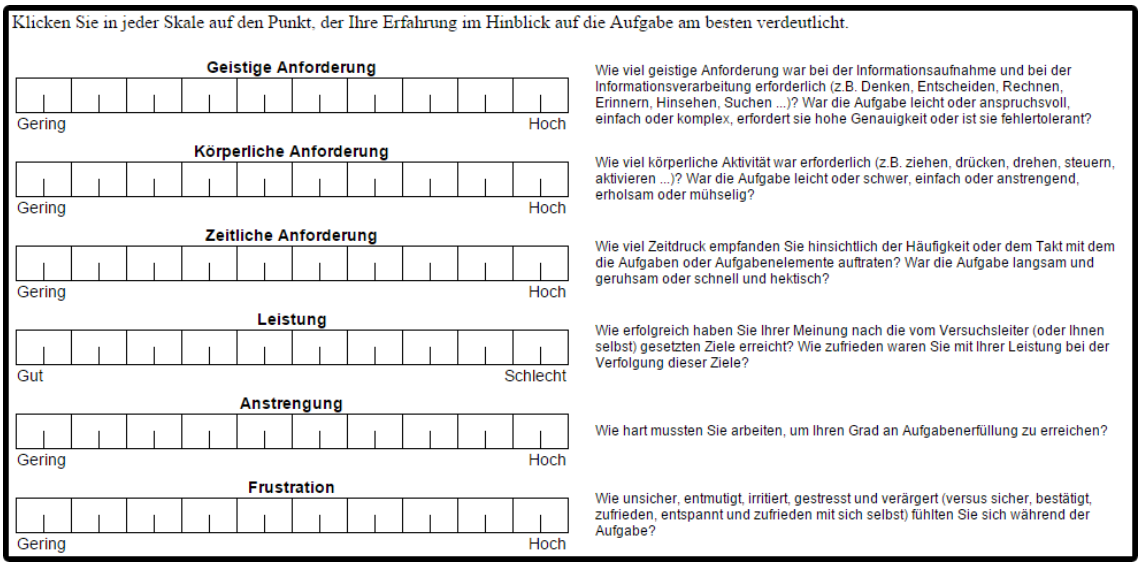
Vinson, N. G. (1999). Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 278–285. ACM.

Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240.

Table of Annexes

A Study Documents

A.1 TLX Questionnaire



A.2 TPI Questionnaire

TPI-Questionnaire PZ907

* Erforderlich

How were the possibilities to coordinate actions between you and your partner? *

	1	2	3	4	5	6	7	
Very bad	○	○	○	○	○	○	○	Very good

How were the possibilities to communicate spatial information with your partner? *

	1	2	3	4	5	6	7	
Very bad	○	○	○	○	○	○	○	Very good

How often did you have the sensation that people you saw/heard could also see/hear you? *

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

To what extent did you feel you could interact with the person or people you saw/heard? *

	1	2	3	4	5	6	7	
None	○	○	○	○	○	○	○	Very much

How much did it seem as if you and the people you saw/heard both left the places where you were and went to a new place? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

How much did it seem as if you and the people you saw/heard were together in the same place? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

1

How often did it feel as if someone you saw/heard in the environment was talking directly to you? *

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often did you want to or did you make eye-contact with someone you saw/heard? *

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

Seeing and hearing a person through a medium constitutes an interaction with him or her. How much control over the interaction with the person or people you saw/heard did you feel that you had? *

	1	2	3	4	5	6	7	
None	○	○	○	○	○	○	○	Very much

For each of the pairs of words below, please circle the number that best describes your evaluation of the media experience. *

	1	2	3	4	5	6	7	
Impersonal	○	○	○	○	○	○	○	Personal

	1	2	3	4	5	6	7	
Unresponsive	○	○	○	○	○	○	○	Responsive

	1	2	3	4	5	6	7	
Unsociable	○	○	○	○	○	○	○	Sociable

	1	2	3	4	5	6	7	
Unemotional	○	○	○	○	○	○	○	Emotional

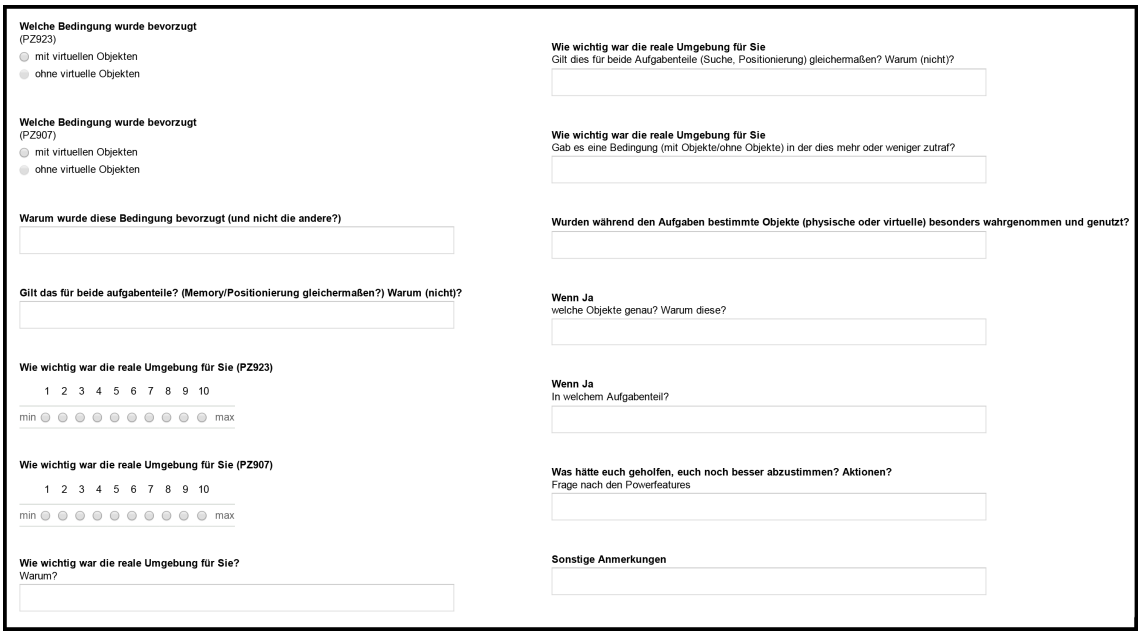
	1	2	3	4	5	6	7	
Insensitive	○	○	○	○	○	○	○	Sensitive

	1	2	3	4	5	6	7	
Remote	○	○	○	○	○	○	○	Immediate

	1	2	3	4	5	6	7	
Dead	○	○	○	○	○	○	○	Lively

2

A.3 Interview Structure



Welche Bedingung wurde bevorzugt
(PZ923)
- mit virtuellen Objekten
- ohne virtuelle Objekten

Welche Bedingung wurde bevorzugt
(PZ907)
- mit virtuellen Objekten
- ohne virtuelle Objekten

Warum wurde diese Bedingung bevorzugt (und nicht die andere?)

Gilt das für beide aufgabenteile? (Memory/Positionierung gleichermaßen?) Warum (nicht)?

Wie wichtig war die reale Umgebung für Sie (PZ923)

1 2 3 4 5 6 7 8 9 10

min max

Wie wichtig war die reale Umgebung für Sie (PZ907)

1 2 3 4 5 6 7 8 9 10

min max

Wie wichtig war die reale Umgebung für Sie?
Warum?

Wie wichtig war die reale Umgebung für Sie
Gilt dies für beide Aufgabenteile (Suche, Positionierung) gleichermaßen? Warum (nicht)?

Wie wichtig war die reale Umgebung für Sie
Gab es eine Bedingung (mit Objekte/ohne Objekte) in der dies mehr oder weniger zutraf?

Wurden während den Aufgaben bestimmte Objekte (physische oder virtuelle) besonders wahrgenommen und genutzt?

Wenn Ja
welche Objekte genau? Warum diese?

Wenn Ja
In welchem Aufgabenteil?

Was hätte euch geholfen, euch noch besser abzustimmen? Aktionen?
Frage nach den Powerfeatures

Sonstige Anmerkungen

A.4 Parameter Settings - Complete Study

dyads	first half				second half		
1	C	K	!B		!C	!K	B
2	C	K	B		!C	!K	!B
3	C	!K	B		!C	K	!B
4	C	!K	!B		!C	K	B
5	!C	!K	!B		C	K	B
6	!C	!K	B		C	K	!B
7	!C	K	B		C	!K	!B
8	!C	K	!B		C	!K	B
9	C	K	!B		!C	!K	B
10	C	K	B		!C	!K	!B
11	C	!K	B		!C	K	!B
12	C	!K	!B		!C	K	B
13	!C	!K	!B		C	K	B
14	!C	!K	B		C	K	!B
15	!C	K	B		C	!K	!B
16	!C	K	!B		C	!K	B

C: With Cues	K: Coordinate Set 1	B: Texture Set 1	!C: Without Cues	!K: Coordinate Set 2	!B: Texture Set 2

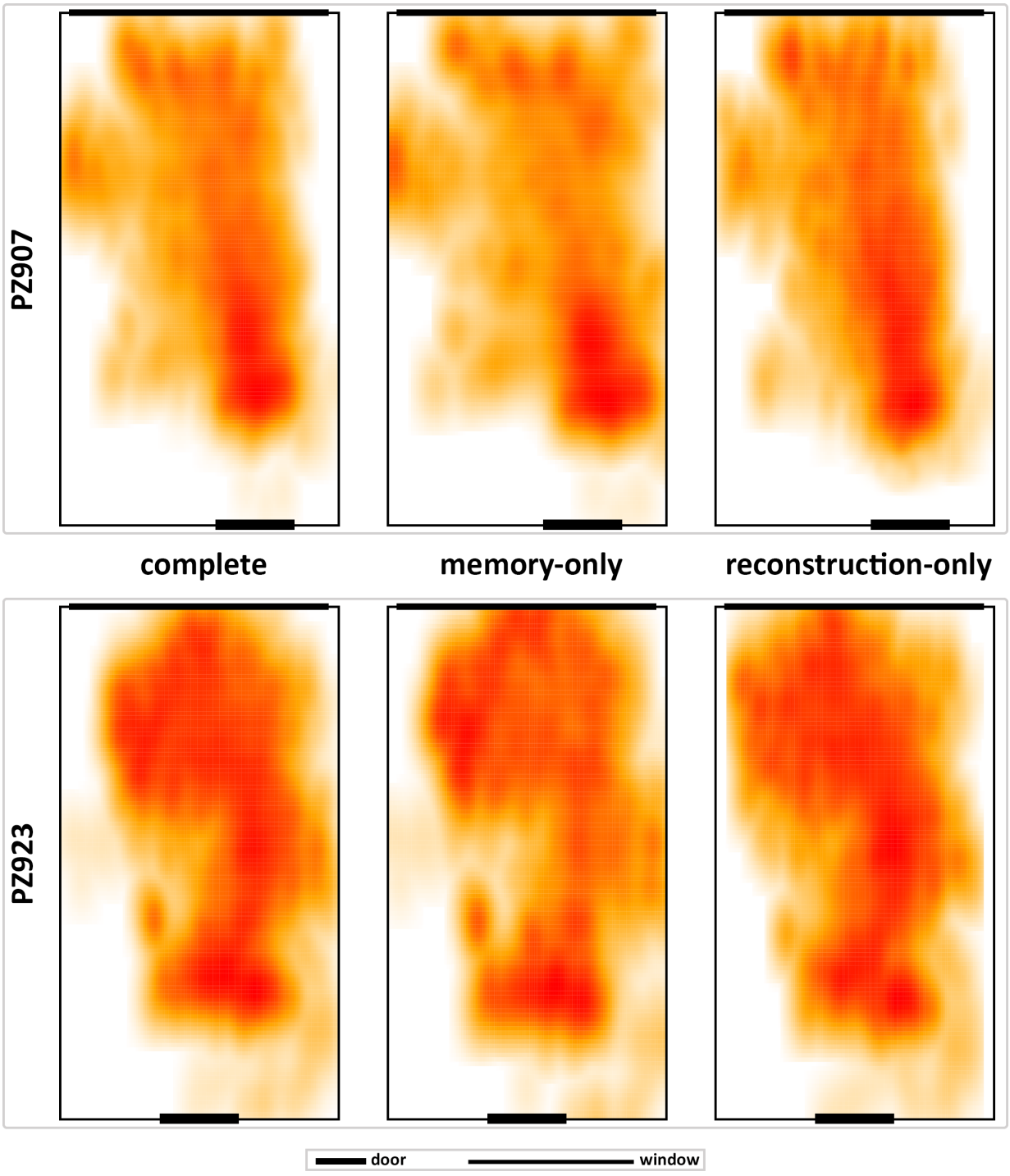
B Evaluation Documents

B.1 Evaluation Video & Audio Footage

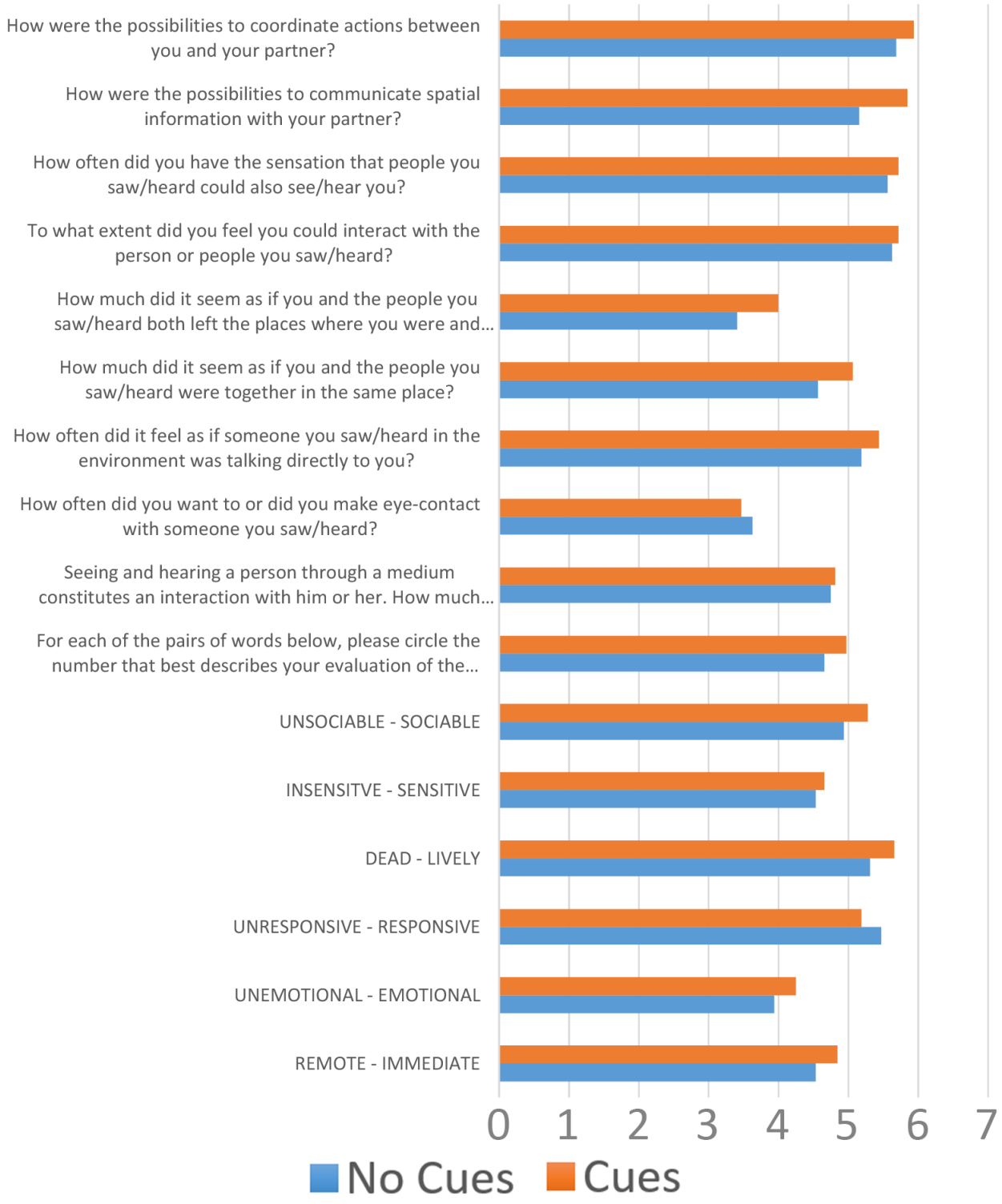
		GRUPPE_____							
	Durchlauf	1		2		3		4	
G	Zimmer	PZ907	PZ923	PZ907	PZ923	PZ907	PZ923	PZ907	PZ923
G	Cues								
G	Koordinaten								
G	Bilder								
R	Selbst (neben, über, rechts von, ... MIR)								
R	Gesten (Visuelle Hinweise / Zeigen (Virtuell mit Box / Frustum))								
R	Mitspieler (neben, über, rechts von, ... DIR)								
R	Boxen (links von, über, neben, vor, ,,, der BOX)								
R	virt. Objekte (links von, über, neben, vor, ... dem virt OBJEKT)								
R	reale Objekte (links von, über, neben, vor, ... dem realen OBJEKT)								
R	Räumlich (weiter Richtung Tür, vorne (Bezug Fenster))								
R	Relative Referenz zu Objekt aus Sicht des anderen Mitspielers								
R	Relative Referenz auf Versuchsleiter								

A	Selbst (bei MIR)								
	Mitspieler (bei DIR)								
	Boxen (bei, ,,, der BOX)								
	virt. Objekte (bei, ... dem virt. OBJEKT)								
	reale Objekte (bei, ... dem realen OBJEKT)								
	Räumlich (beim Fenster, bei der Tür)								
	Absolute Referenz auf Versuchsleiter								
S	Deiktische Referenz (hier, da)								
	Referenzen auf Reale Objekte im ANDEREN Raum								
	Referenz auf Position Zwischen Zwei Boxen/Virtuellen Objekten								
	Lösungsstrategie								
	Selbstgespräch								
	Sonstige Anmerkungen								

B.2 Heat Map: Player Positions



B.3 TPI: Average Values - Object Identification Task



C Digital Copy