Mediated Attention with Multimodal Augmented Reality

Angelika Dierker
CITEC, Faculty of Technology
Bielefeld University, Germany
adierker@techfak.uni-bielefeld.de

Christian Mertes
CITEC, Faculty of Technology
Bielefeld University, Germany
cmertes@techfak.uni-bielefeld.de

Thomas Hermann
CITEC, Faculty of Technology
Bielefeld University, Germany
thermann@techfak.uni-bielefeld.de

Marc Hanheide
School of Computer Science
University of Birmingham, UK
m.hanheide@cs.bham.ac.uk

Gerhard Sagerer
CITEC, Faculty of Technology
Bielefeld University, Germany
sagerer@techfak.uni-bielefeld.de

ABSTRACT

We present an Augmented Reality (AR) system to support collaborative tasks in a shared real-world interaction space by facilitating joint attention. The users are assisted by information about their interaction partner's field of view both visually and acoustically. In our study, the audiovisual improvements are compared with an AR system without these support mechanisms in terms of the participants' reaction times and error rates. The participants performed a simple object-choice task we call the *gaze game* to ensure controlled experimental conditions. Additionally, we asked the subjects to fill in a questionnaire to gain subjective feedback from them. We were able to show an improvement for both dependent variables as well as positive feedback for the visual augmentation in the questionnaire.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces (D.2.2, H.1.2, I.3.6)—*Interaction styles*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Computer-supported cooperative work*; H.1.2 [**Models and Principles**]: User/ Machine Systems—*Human factors*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

General Terms

Experimentation, Human Factors

Keywords

Multimodal, joint attention, field of view, CSCW, augmented reality, mediated attention, artificial communication channels, collaboration

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.

Figure 1: Equipment of one subject. The subjects wear AR goggles, headphones and cloth to ensure that they do not bypass the goggles.

1. INTRODUCTION

Augmented Reality (AR) as an interface for closer human-machine interaction and as assistive technology is studied for quite some years dating back to the 90s. The idea to augment the human senses using technology allows AR systems to present additional information in a very context-related way. Hence, many AR systems focus on aspects of appropriate *presentation techniques* and *context analysis*. Most application scenarios exploiting AR consider the use-case of knowledge-based assistance, as for instance in an early work on maintenance and repair tasks [3] or for visual illustration, e. g. in the architecture of Wisneski et al. [14]. A typical example for context-related AR are interactive tourist guides as presented by [12] based on the "Studierstube" AR environment.

However, AR has mostly been discussed and studied as a rather individual assistive technology, at least for *wearable* AR systems that augment the field of view of a wearer of head-mounted displays. In other words, such AR systems can be seen as an extension of human's cognitive abilities, for instance as an external but easily accessible memory or as a means to guide attention to relevant aspects of the seen.

During the years, such wearable AR systems have indeed advanced and their usability in different assembly tasks has been demonstrated and proven [13]. Although there already exist several systems that enable a *collaborative AR* which allows multiple users to share a common mixed reality, such wearable AR systems have a significant drawback: Although they introduce virtual stimuli and information to augment

the real world, they also reduce the user's perception of this world. Despite new achievements in hardware design such as high quality see-through displays allowing less intrusive embedding of virtual content in the real world, all devices at least partially shroud the eye area of the wearer. Although this is usually no direct problem for the wearer him- or herself, it has a negative effect in human-human collaboration scenarios. In collaborations, humans can benefit to a significant extent from direct visual eye-contact. Several interaction-relevant cues require eye-contact between the interactants; a most prominent one considered in this paper is *mutual attention*. Assume the following scenario of collaboration:

> Angie and Chris both inspect a car's engine. Angie, as an expert, tries to explain to Chris the operation principles of the engine to allow him to understand the cause of a malfunction. Both wear AR systems that individually augment their field of view with textual information overlays regarding the motor operation characteristics. So the systems convey additional knowledge to both of them individually. Chris, however, has problems to follow Angie's explanations because it is hard for him as a novice to identify individual parts Angie refers to such as the ignition coil or the spark plug. He would require either a clear deictic gesture of Angie to understand her reference or – even simpler and more fluent – look at her face and eyes to understand what she might be looking at.

This scenario indicates what we are targeting at in this paper. We present an AR system designed to particularly *support* collaborative tasks in shared spaces by establishing an AR-based coupling between the two to facilitate joint attention. In the study we present in Section 3, two users have to jointly solve a well-structured task with regard to the manipulation of virtual objects in a real-world table setting. Both are equipped with wearable AR systems and we explicitly enhance their interaction abilities with a multimodal mediation of their mutual foci of (visual) attention. In previous work, for instance in Kalkofen et al. [5], AR techniques have already been employed to guide a single user's attention in a context-aware manner. In our work instead, we *closely couple* two AR systems, exploiting one's field of view as contextual information for the augmentation of the other.

At this point, it shall be noted that although AR is basically multimodal, the term "augmented reality" is mostly interpreted as *visual* augmentation by the scientific community. However, in this paper we will discuss AR as an audio-visual technique to convey attention-related information to the interactants in real-time, hence enabling them to be closely coupled by means of an AR interface. In the paper we will not only present how our prior system [10] is extended, but also show for the first time that it significantly improves the collaboration both quantitatively and qualitatively.

2. MEDIATED ATTENTION

Attention can be defined as the mechanism for the goal-specific allocation of limited perceptual resources. Mechanisms of attention are well studied in the visual domain, for instance using eye-tracking methods in controlled experiments [7]. In the context of human-human and mediated cooperation, attention touches different aspects: (a) the mechanisms used by interlocutors to allocate their perceptual resources (e. g. focus on a visual region of interest or attending a certain signal stream in the soundscape), (b) the methods and signals used by the cooperating users to draw, shape or direct the other's attention. It may be assumed that joint attention – the intentional coupling of attention – supports, or even enables cooperation, particularly in the case where the interactants' internal representations differ regarding their current context.

Certainly, biological evolution produced a multitude of strategies to best employ limited perceptual resources, particularly in cooperation. For instance, we are capable of interpreting other people's focus of attention from observing their head orientation, gaze, and often the body posture and arm gestures. Pointing and other indexical gestures are commonly used to guide another person's focus of attention and we often are not aware of the complexity of these mechanisms since they are subconsciously and routinely used.

In light of these mechanisms, two research questions arise when we aim to support cooperation by means of technical systems: (1) how are our natural communication channels disturbed or derogated in case of an indirection through technological means (e. g. by intercepting direct visual eye-contact with head-mounted systems) and what effect has this on cooperation? And (2) how and under which circumstances can new forms of technical mediation actively contribute to support joint attention, thereby forming a kind of *artificial communication channel*? For example, can signals be displayed that accelerate the process of joining attention?

We are investigating both questions with our AR system depicted in Figure 1 and focus here on question (2). In general, since natural communication channels are very good and humans even coevolved phylogenetically with them as stated above, technical systems need to find niches where they can contribute. We currently see two such niches: (1) the compensation of disadvantages the technology introduced itself if it is necessary to use it, and (2) exploiting the attention bottleneck, i. e. retaining and providing data that the user in principle would have been able to perceive but actually did not because the attention was concentrated somewhere else. These situations of divided attention should become even more frequent in multi-user scenarios. Therefore, multi-user capabilities might be a useful extension of our system.

In the following we present two new augmentation strategies that use coupled AR systems in a shared coordinate system: firstly, the visual augmentation of elements in the other's field of view, and secondly, the auditory augmentation by means of sonification (cf. Section 2.2) depending on whether objects are visible by the partner.

2.1 Vision-Mediated Attention

Visual augmentations can manipulate the users' attention in a variety of ways. For selecting the best method it would be necessary to fully understand the mechanisms that guide the user's attention in visual exploration. Since the interplay of these mechanisms differs from situation to situation, a single simple answer may not exist. However, by categorizing attention as a mixture of subconscious and conscious processes (such as the explicit searching for certain patterns), we can at least suggest some visual augmentation types.

For instance, a localized adaptation of saliency (e. g. local image filters such as changing the contrast or brightness, or applying low- or high-pass filters) will change the underlying

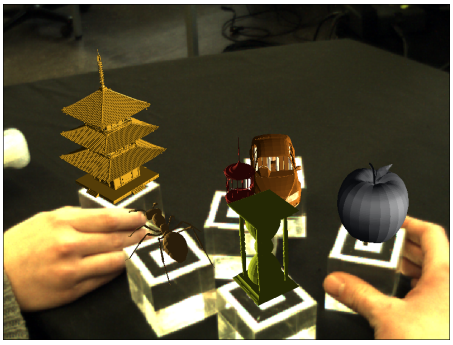


Figure 2: Screenshot of the ARToolKit virtual objects and the visual highlighting in action. The cage is in the center of the partner's field of view (red), the apple is not seen at all (gray), and the other objects are seen more or less peripherally (orange-yellow-green).

basis for our existing visual attentional processes and lead, for instance, to quicker (or slower) detection of the thereby pronounced (or obscured) object. Suchlike techniques that highlight or augment a specific area in the field of view are often referred to as *magic lenses* and have been successfully employed to guide attention in AR [8]. Alternatively, temporal patterns such as blinking at certain locations are a strong and salient cue to guide the eyes (or in our case the head), yet such elements need to be used carefully since they may disturb more than they help because they are also strong distractors from otherwise relevant information [11]. Another type of augmentation effect would be the localized magnification of regions, using a local nonlinear distortion (like a fovea) which lets highlighted image regions cover more space in the field of view of the user.

Such highlightings are interesting – however, they are computationally expensive and radical in the way they break with the user's normal visual perception. We therefore use a more basic yet effective form of visual augmentations that are more easily implemented, offer good control and a good experimental examination of how mediated attention affects AR-based cooperation: we augment gray-colored virtual objects on top of physical objects using the ARToolKit marker system [6] and control the color (hue) of the virtual objects for one user according to the object locations in the field of view of the *other* user and vice versa. More precisely, the color changes from yellow (peripheral) to red (in the center of the partner's field of view). This coupling between users can of course be implemented unidirectionally in asymmetric cooperation tasks (cf. Figure 2). To enable the system to be useful in situations of temporarily divided attention, i. e. one user looks at an object a moment after his partner has looked away, we have the color highlighting fade in and out with configurable times and envelopes. The fade-in is useful to prevent a quick glance or a sweep over an object from letting it after-glow as if the focus of attention had rested on it for a substantial amount of time.

Other types of vision-mediated attention, such as the direct indication of the field of view as vision cone or projection onto a surface are conceivable and might be intuitive. However, they were not yet used in the presented study although the projection method has already been suggested and implemented in our previous work [10]. A comparison of different ways to enhance cooperation by such attention mediation is subject of our ongoing research and will be reported elsewhere.

2.2 Sonification-Mediated Attention

Sonification, the non-speech auditory display of information by using sound is a young and rapidly growing alternative and complement to visualization.[1] From everyday experience we are quite familiar with the fact that sound often directs our attention, e. g. towards approaching cars when we cross a road, towards somebody calling us, or towards the boiling water kettle in our kitchen. In these situations, our ears guide our eyes and therefore we regard it as an interesting approach to explore the possibilities to use sound to couple the attention of cooperating users.

For this novel approach of using sonification for interaction support we see various possible methods and thus far have only scratched the surface of what we expect to become a promising field in auditory AR technologies. To implement our ideas of mediated attention via sonification as an additional channel to the visual highlighting explained above, we continue with the same approach of using objects (tracked via ARToolKit) which may be in the interlocutors' field of view (cf. Section 2.1). The simplest sonification strategy is to play a short sound event (e. g. a click sound) whenever an object enters the partner's field of view. Using different sounds for entering and leaving (e. g. 'tick' – 'tock'), each user receives some awareness about the overall head activity without the need to look at the partner's head.

So for this paper we use sonification only as a marginal information stream, displaying by sound whenever objects enter or leave the partner's field of view. This is because at this stage we are primarily interested in the overall influence of mediated attention on performance in cooperation. In our ongoing work we will consider sonification as an artificial communication channel in more detail though, as semantically more meaningful sonifications are easily conceived. We are currently working on the implementation of two examples: One uses the objects in the interaction partner's field of view as does the current sonification. Each virtual object, however, could then constantly emit a sound that seems to originate from its actual position while the amplitude of the sound correlates with the distance to the interaction partner's field of view. Using a sound that is easily associated with the particular object (like a car engine for a car) could allow users to draw upon more of their already existing knowledge and skills. Alternatively, the sonification might map different attributes of the head direction to a sound which might intensify with movement or proximity of the interaction partners' fields of view [9].

2.3 Problem and Hypotheses

As we argued in Section 1, AR systems on the one hand augment the real world but on the other hand they also reduce the user's perception of this world (e. g. by shrouding the wearer's eye area). In order to oppose this, we decided to re-enhance the interaction by giving the participants the audiovisual augmentations explained above.

[1]The main community is the ICAD (see `www.icad.org`, seen May 2009). A comprehensive definition of sonification is given in [4] (see also `www.sonification.de`, seen May 2009).

We assume that these audiovisual augmentations using the AR goggles will only enhance interaction when compared to a setting also using AR goggles for the simple reason that currently these devices are too disadvantageous in terms of lag, resolution, field of view, weight, shutter time, monoscopic vision and dynamic range. Compared to a setting where both participants wear no AR devices we therefore do expect the general problems of AR to outweigh the benefits of the augmentations to enhance interaction. However, our point is that we are able to at least partially compensate some of the disadvantages of AR devices (mainly the reduced field of view and the hampered head movements) with this technique.

Our hypotheses are (1) that the participants have a *lower error rate* in the condition using the audiovisual augmentations compared to the condition without them, and (2) similarly that they exhibit a *shorter reaction time* in the condition with both augmentations.

3. METHOD

In the following, we give an overview of the experimental setup for the study. Beginning with a description of the tested conditions and the measurements we proceed with information about the sample, the stimuli and the procedure of the experiment.

To test if the augmentations actually improve the interaction using AR via HMDs, we chose to compare two experimental conditions: the "highlighting on" condition where both the visual and the auditory augmentations were provided and the "highlighting off" condition where neither visual nor auditory augmentation was given. The reason not to distinguish between the visual and the auditory augmentations in the experimental conditions was to test every pair of participants in all conditions while still not overburdening the participants with a very long experiment (most people get tired wearing the AR goggles over a longer period of time [1]).

As dependent variables we consult objective as well as subjective criteria to measure the effect of the "highlighting on" condition. Objective dependent variables are reaction time (the time needed to finish the task successfully) and error rate (the percentage of successfully finished tasks). The reaction time is measured using button presses on Wii Remotes[2]. In each phase of the trial the participants have to press the button of the Wii Remote to continue with the trial. Error rates are calculated from offline annotations of the scene camera data.

Additionally, we determine the subjective user experience in a questionnaire with multiple-choice answers. We asked the participants two questions to estimate their subjective usage of the augmentations. The first was "How much did you use the visual augmentations?", the second was "How much did you use the auditory augmentations?". There were four checkboxes for the answers and the scale ranged from "very much" to "not at all" with the intermediate steps not named explicitly. Moreover the participants answered the questions "How helpful did you find the visual augmentations?" and likewise, "How helpful did you find the auditory augmentations?". Here, the scale ranged from "very helpful" to "not helpful at all" in four steps (intermediate steps were not named explicitly). There was a fifth possible answer "disturbing" for these two questions.

[2]cf. `http://www.nintendo.com/wii/what/controllers#remote`

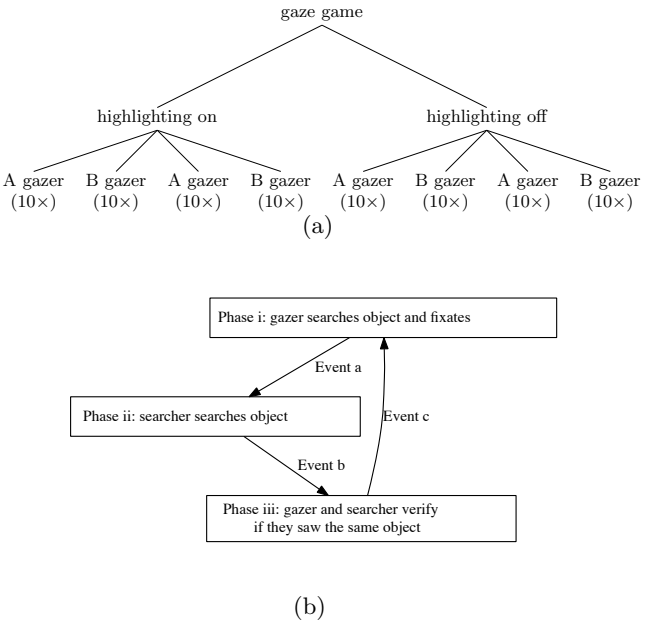


Figure 3: (a) Process flow of the gaze game. The tree is traversed left to right, every leaf being an action the subjects had to take. The leaf nodes also indicate whether subject A or subject B takes the role of the gazer and that ten cycles are to be completed before switching roles. The order of "highlighting on" and "highlighting off" was interchanged for half of the subject pairs. (b) Phases of one cycle of the trials which are started/ended by Wii button events (see text for explanation).

3.1 Scenario: The "Gaze Game"

To evaluate the effectiveness of the augmentations described in Section 2 and to test our hypothesis, we used a task that we call the *gaze game*. In this, a pair of subjects plays cooperatively, sitting next to each other at a table. Each of them is alternatingly allocated one of two roles: one player is the *gazer* and the other one is the *searcher*. Each game cycle (see Figure 3(b)) starts with the gazer getting displayed one of the virtual objects in the corner of the head-mounted display. Subsequently, he or she has to find the corresponding object on the table and fixate it (phase i). When this is accomplished, the gazer presses a Wii Remote button to trigger a vibration in the searcher's Wii Remote for whom this is the signal to find the same object as fast as possible (phase ii), indicating a presumed success with another button press. No speaking or gesturing is allowed up to this point in each cycle. The time between these two button presses is taken as a performance measure. The two players are then allowed to speak again and are in fact asked to verify whether they were indeed looking at the same object (phase iii). A final button press by the gazer finishes a cycle, begins a new one and triggers a new object to be displayed. The placement of the objects on the table is randomly changed at that moment, too, to prevent the players from learning the positions by heart. After ten cycles the roles of gazer and searcher are switched. The marker positions at which the objects were displayed are shown in Figure 4.

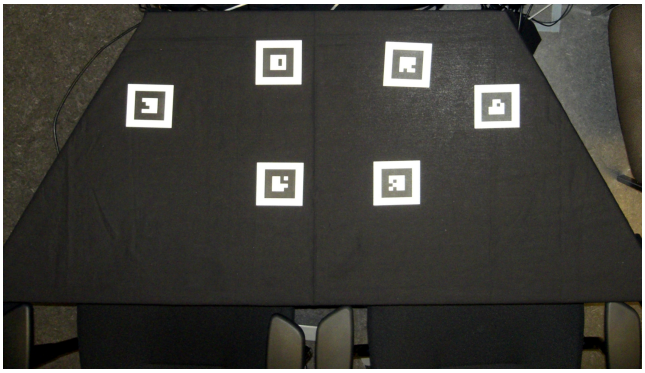


Figure 4: Table used and the arrangement of markers during the gaze game. The long side of the table measured 140 cm, the opposite side 70 cm and the depth was 61 cm.

After two blocks of ten cycles per person, the highlighting condition is changed, so that each pair of subjects plays 40 cycles with augmentations and 40 cycles without. Specified by the order of their arrival, the pairs of subjects began alternatingly with augmentations and without augmentations respectively.

3.2 Procedure for the Study and Sample

The trials for the study consisted of three parts: In the first part, the participants were equipped with the devices and were given time to explore and adapt to the AR system. This was supported by letting them sort and group the objects regarding several criteria. The subjects were asked whether they felt comfortable with the system before proceeding. The visual and auditory augmentations were switched on during this phase for all subjects to give everyone the opportunity to get used to the full system. The second part was the gaze game task we explained above. The subjects wore the AR goggles and headphones the whole time. Additionally, their eyes were laterally shielded to prevent them from bypassing the goggles as some subjects did in a preliminary study. During the whole task, time measurements of Wii Remote button presses were taken. Finally, subsequent to the gaze game, the participants were asked to put down the equipment and to fill in a questionnaire.

For this study, we tested 13 pairs of subjects. Unfortunately, we had to leave out the data from two pairs of subjects later. Of one pair, because they misunderstood the "highlighting off" condition in the gaze game task by guessing instead of looking at their partner and estimating the partner's gaze direction. The other pair could not finish the task due to time limitations. The remaining 22 subjects were 11 male and 11 female. Their age ranged between 18 and 28 years and the mean age was 21.8 years. The majority of the sample were students. The participants were asked to schedule their appointment using an online tool. Thus, we had no influence on the composition of the subject pairs. Nine out of eleven pairs had different gender, the two remaining pairs had the same gender. Seven pairs did know each other beforehand. Six of the pairs of subjects began with augmentations, five of them without. Each subject played only in a single pair. The duration of the gaze game task varied from 16 to 26 minutes with an average of 21 minutes. All participants were paid a fee for their participation.

4. RESULTS

From the gaze game part of the study, we measured the reaction time and computed error rates. We annotated all scene camera videos offline according to the participants' success in the task. Thereby, we left out all trials that could not be rated (because of disturbances or technical problems) for computing both the reaction times and the error rates. The performance results are presented in this section as well as the questionnaire results.

4.1 Reaction Time

In the gaze game task, the main goal to achieve was to get the right object not to be quick by all means. Thus, we exclusively considered the successful cycles for the reaction time because in the unsuccessful ones the searchers pressed the button before they had actually found the right object. Time measurements were taken from logged timestamps of Wii button presses: we computed the difference of the button press constituting Event a and the one being Event b (see Figure 3(b)). The mean search time[3] for the "highlighting on" condition is 2.56 ± 0.98 seconds whereas the mean search time for the "highlighting off" condition is 4.16 ± 1.97 seconds. The comparison of the means for reaction time with paired two-sample t-test[4] showed statistical significance [$T(10) = -2.5$; $p = 0.03$]. As visualization, a boxplot is shown in Figure 5 for both conditions. The boxes span between the lower and the upper quartile, the horizontal lines are the medians whereas the thick black bars depict the mean values. Whiskers show the minimal and maximal search times. For each subject pair the conditions are given in the correct order (first condition on the left). The "highlighting on" condition is represented with red boxes whereas gray boxes represent the "highlighting off" condition. The last two boxes show the overall search times for both conditions (yellow (left) for the "highlighting on" condition and blue (right) for the "highlighting off" condition). For this plot only the cycles with correct outcome were considered. The means for the overall search times visualize our previous results: the mean for the "highlighting on" condition is lower than for the "highlighting off" condition. Considering the means for the subject pairs, we can see that except for subject pair 6, all means of the subject pairs are lower for the "highlighting on" condition than for the "highlighting off" condition.

4.2 Error Rates

For the "highlighting on" condition, the subjects made an average of $2.72 \pm 3.41\%$ errors whereas for the "highlighting off" condition the average error rate was $36.86 \pm 15.88\%$. Strong significance is found by a paired two-tailed t-test[5] from the error rates [$T(10) = -6.39$; $p < 0.01$]. A comparison of both graphs in Figure 6 visualizes this result.

[3]notation: arithmetic mean $\pm$ standard deviation

[4]Homogeneity of variances was shown by f-test [$F(10, 10) = 0.25$; $p = 0.04$], both samples have underlying normal populations.

[5]Homogeneity of variances was shown by f-test [$F(10, 10) = 0.05$; $p < 0.01$], both samples have underlying normal populations.

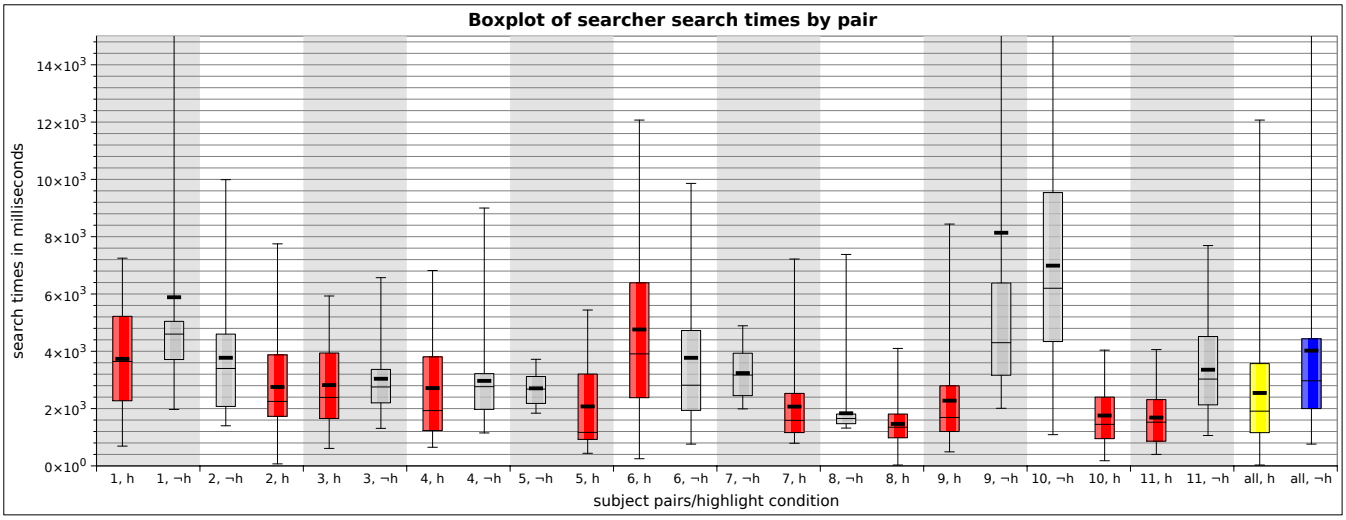


Figure 5: Boxplot of the searcher's search times. Boxes show the interquartile range and the medians (horizontal lines), thick black bars depict the mean values. Whiskers show minimal and maximal search times. For each subject pair, the conditions are given in the correct order (first condition left). Red (darker) boxes represent the "highlighting on" condition, gray (brighter) boxes represent the "highlighting off" condition. The last two boxes show the overall search times for both conditions (yellow/brighter for the "highlighting on" condition and blue/darker for the "highlighting off" condition). Only the successful cycles were considered.

4.3 Questionnaire

For the question "How much did you use the visual/auditory augmentations?" the questionnaire results are shown in Figure 7(a). We found that most participants rated their usage of the visual augmentation very high while they rated their usage of the auditory augmentation very low. Thus, there is a clear difference between the rated usage of the visual augmentations compared to the auditory augmentations.

For the question "How helpful did you find the visual/ auditory augmentations?" the answers are presented in Figure 7(b). Most participants found the visual augmentations helpful. However, 16 participants rated the helpfulness of the auditory augmentations on the negative half of the scale and 4 subjects found them even disturbing. Thus, there is a clear difference between the valuation of helpfulness for the visual versus the auditory augmentations.

5. DISCUSSION

Our hypotheses were (1) that the participants have a *lower error rate* in the "highlighting on" condition compared to the "highlighting off" condition and (2) similarly that they exhibit a *shorter reaction time* in the "highlighting on" condition.

For the reaction time, we found a significant difference between both highlighting conditions and thereby support our hypothesis 2. Moreover, only for one subject pair, the mean in the reaction time for "highlighting on" condition is lower compared to the "highlighting off" condition. We suppose that this is due to a lack of faith of this subject pair in the reliability of the highlighting which resulted in multiple verifications of the assumed view direction each time before communicating the decision. Nevertheless, we can conclude that the audiovisual augmentations cause a lower search time for the searcher in general.

Moreover, the measured error rates support our hypothesis 1 because they are significantly lower for the "highlighting on" condition. This means that in sum, both augmentations (visual and auditory) together induce a faster and less error-prone task completion.

In a pre-study [9], we found similar error rates for the "highlighting on" condition but different ones for the "highlighting off" condition compared to these results. When we realized in the pre-study that the task is too easy to measure effects in the error rates, we altered several parameters of the task. We propose the following four changes we made to the task between the pre-study and the study presented in this paper as possible explanations for this difference: Firstly, the gazers are no longer allowed to choose an object by themselves because we present a random object to them. Secondly, the positions of the objects on the table are shuffled before each cycle so that the searcher can not learn the objects by heart. Thirdly, we increased the number of possible objects from five to six. Fourthly, we prevent the participants from looking past the goggles by blinders. To explain that the task became more difficult yet the error rate of the "highlighting on" condition was not affected, we suggest the improved visual augmentation. While the highlighting was simply yellow for all objects in the partners' field of view in the pre-study, the highlighting in this study is a color gradient from the center of the partners' field of view to the outer region. We consider a combination of both randomization techniques and the increased number of possible search objects to be the crucial factor for the different error rates for the "highlighting off" conditions of both studies. Nevertheless, there is no proof for this hypothesis yet. This effect might as well has been caused by the influence of the blinders which should be investigated in a subsequent study.

Concerning the subjective user experience measured by the questionnaire, we found a pronounced difference between the

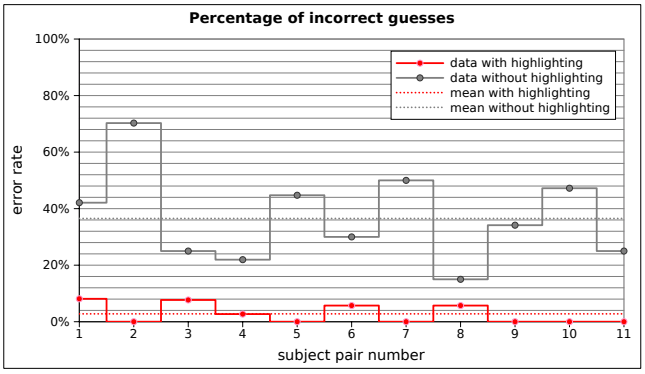
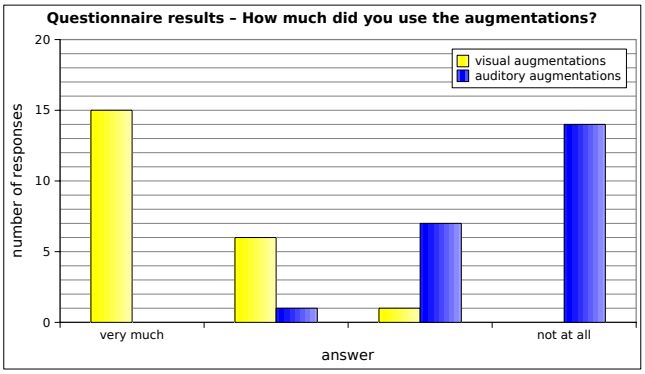
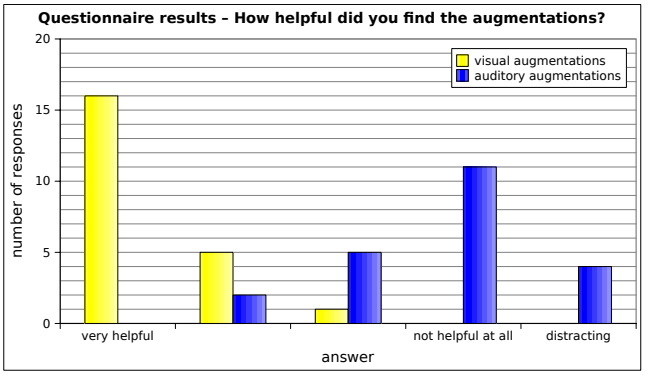


Figure 6: Error rates for both tested conditions. The red dots (on the lower curve) belong to the "highlighting on" condition whereas the gray dots (on the upper curve) belong to the condition where the highlighting was switched off. Note that the connection of the dots does not indicate interim values. Dotted horizontal lines specify the average values for both conditions.



(a)



(b)

Figure 7: Answers from questionnaire. The intermediate steps were not named explicitly.

augmentation modalities. The participants found the visual highlighting much more helpful than the auditory highlighting. Some even rated the sonification distracting. Similarly, they rated their usage of the visual augmentations much higher than their usage of the auditory augmentations. We suggest three possible reasons for this: Firstly, there is much less information conveyed by the auditory augmentations than by the visual ones. While the visual augmentations make clear which object is being looked at and how centrally, the auditory augmentations' main potential function is that of an activity monitor. Even this was impeded by the fact that even moderately fast head movements caused the image to be blurred to the point where the markers were unrecognizable. More sophisticated sonifications might therefore still be promising (cf. Section 2.2 and [9, 10]). Secondly, the visual modality is more commonly used for joining attention in everyday communication than the auditory one. Therefore, there might be a training effect which could be shown by longitudinal studies. Thirdly, auditory cues could work more on a subconscious level than visual ones. This should be shown by separating the highlighting modalities in consecutive studies.

6. CONCLUSION AND OUTLOOK

In this work, we presented an AR system to support collaborative tasks by facilitating joint attention. Two collaborators are equipped with wearable AR devices which are closely coupled so that the virtual objects lying in the partners' field of view are highlighted in the users' AR gear. This highlighting is both visual and acoustical. Visual, by changing the color of those objects according to their spacial distance from the center of the field of view. Acoustically, an auditory highlighting is provided which simply indicates appearance and disappearance of objects with two different sounds. Our hypothesis was that these highlightings would reduce reaction times as well as error rates. We tested both augmentation modalities together and contrasted them with a condition in which neither augmentation method was provided. For controlled experimental conditions we used the *gaze game*, a simple object-choice task. With this, we measured reaction times and error rates. Additional to the objective measurements, we asked the participants to fill in a questionnaire to obtain subjective ratings.

The results for the reaction times and the error rates show a clear support of our hypotheses and thus that our augmentations can improve AR-based cooperation. In the questionnaire results, we found a clear difference between auditory augmentation and the visual augmentation. Most participants found the visual highlighting helpful while the auditory augmentation was not rated to be helpful. Whether people still manage to incorporate these or other attention focus indicators when facing a more demanding interaction situation should be investigated in future studies. We also did not compare users of our AR system with subjects unencumbered by any AR devices as we do not expect the presented techniques to outweigh the disadvantages of current AR hardware.

It is therefore currently very likely that artificial communication channels as presented in this work could at most be applied where AR is already used, as for cooperation in laboratories or factories. However, with head-mounted displays and wearable computers becoming less obtrusive, such signals might even have the potential to support nat-

ural interactions. Despite the unsatisfying evaluation of sonification in this study, it has the advantage to keep the eyes free and may be convenient with acoustically transparent headphones. Furthermore, while headphones already are ubiquitous, it will take time until the everyday user is equipped with lightweight and practicable AR goggles. Thus, the auditory channel may be faster to find its way into application. Therefore, further examination of this modality might be particularly rewarding.

Apart from this, our ongoing research is to explore and investigate ways and prerequisites of mediating attention, to further develop artificial communication channels [9] and to support and to extend the capabilities of our AR system in order to be able to conduct studies that help to gain a better understanding of human's communication abilities [2].

7. ACKNOWLEDGMENTS

This work has partially been supported by the Collaborative Research Center (SFB) 673 *Alignment in Communication* and the Center of Excellence for Cognitive Interaction Technology (CITEC). Both are funded by the German Research Foundation (DFG).

8. REFERENCES

[1] K. Arthur. *Effects of field of view on performance with head-mounted displays*. PhD thesis, University of North Carolina, 2000.

[2] A. Dierker, T. Bovermann, M. Hanheide, T. Hermann, and G. Sagerer. A multimodal augmented reality system for alignment research. In *Proceedings of the 13th International Conference on Human-Computer Interaction*, pages 422–426, San Diego, USA, July 2009.

[3] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based augmented reality. *Commun. ACM*, 36(7):53–62, 1993.

[4] T. Hermann. Taxonomy and definitions for sonification and auditory display. In B. Katz, editor, *Proc. 14th Int. Conf. Auditory Display (ICAD 2008)*, pages 1–8, Paris, France, June 2008.

[5] D. Kalkofen, E. Mendez, and D. Schmalstieg. Interactive focus and context visualization for augmented reality. *Mixed and Augmented Reality, IEEE/ACM International Symposium on*, 0:1–10, 2007.

[6] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, 99:85–94, 1999.

[7] H. Koesling. *Visual Perception of Location, Orientation and Length: An Eye-Movement Approach*. PhD thesis, Bielefeld University, 2003.

[8] E. Mendez, D. Kalkofen, and D. Schmalstieg. Interactive context-driven visualization tools for augmented reality. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality ISMAR 2006*, pages 209–218, 2006.

[9] C. Mertes. Multimodal augmented reality to enhance human communication. Master's thesis, Bielefeld University, Aug. 2008.

[10] C. Mertes, A. Dierker, T. Hermann, M. Hanheide, and G. Sagerer. Enhancing human cooperation with multimodal augmented reality. In *Proceedings of the 13th International Conference on Human-Computer Interaction*, pages 447–451, San Diego, USA, July 2009.

[11] M. Posner and Y. Cohen. Components of visual orienting. *Attention and performance X*, pages 531–556, 1984.

[12] G. Reitmayr and D. Schmalstieg. Collaborative augmented reality for outdoor navigation and information browsing. In *Proc. Symposium Location Based Services and TeleCartography*, 2004.

[13] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 73–80, New York, NY, USA, 2003. ACM.

[14] C. Wisneski, H. Ishii, A. Dahley, M. Gobert, S. Brave, B. Ullmer, and P. Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In N. Streitz, S. Konomi, and H. Burkhardt, editors, *Proc. Int. Workshop on Cooperative Buildings*, pages 22–32, Darmstadt, Germany, Feb. 1998.