An augmented-reality-based scenario for the collaborative construction of an interactive museum

Angelika Dierker, Karola Pitsch, Thomas Hermann
e-mail: adierker@techfak.uni-bielefeld.de
University of Bielefeld, Collaborative Research Centre SFB 673

Abstract

This paper proposes a scenario for the analysis of interaction mediated by AR. Using this scenario (a) we can easily track all objects in space and over time and record who handles each at which moment, (b) we can easily adjust the displayed augmented objects, (c) we can add meta-information next to the objects in the users' visual field, and (d) we can explore truly multimodal interactions, such as allowing users to perceive the soundscape at any location on the plan by interactively mixing the acoustic contributions that the exhibits make. Most importantly, we will be able to control which information will be perceived by which participant, for example, presenting different features of the object to the two participants (small vs. big, silent vs. noisy, etc.), so that we are able to induce potentially problematic situations which will allow us to investigate how participants deal with such non-obvious misinterpretation of the setting.

1 Introduction

More than forty years ago, Sutherland [1968] presented the first head-up display that could be used for superimposing virtual objects on real world images – the Augmented Reality (AR) technique. Since then, numerous contributions to enhance hardware and software have been made. In the past years, the ongoing trend in AR is to develop less obtrusive and thereby less disturbing devices, carried by the vision that AR systems may be as ubiquitous as sunglasses and mobile phones in the future (e.g. Papagiannakis et al. [2008]). Today's applications for AR techniques include tele-presence, remote control, games, supportive systems in industrial production environments and personal assistance systems (e.g. Ulbricht and Schmalstieg [2003], Wrede et al. [2006]).

Apart from these applications, we have suggested to use AR also as a tool for research on human-human interaction [Dierker et al., 2009a]. The proposed system (the Augmented-Reality-enaBled INterception Interface – ARbInI) allows to conduct controlled experiments in task-oriented interaction. The goal is to intercept (monitor, record) and manipulate (augment, enhance, disturb) the user's auditory and visual perception of the world (and thus important signals transmitted during the interaction) in real-time. To achieve this goal, the system consists of microphones, inertial sensors and video see-through goggles. A modular software evaluates and saves the sensor data (speech data, head movements, objects in the field of view) to a database for further (offline) analysis. Moreover, an AR visualisation system embeds 3D graphics on top of visual markers attached to wooden cubes. For the users, these graphics appear as objects solidly connected to the cubes so that they can naturally take them into their hands and inspect them from all sides. These augmentations provide the basis for AR-based object games as controlled stimuli for dyadic cooperative interactions [Mertes et al., 2009]. The aim is to systematically investigate this task-oriented communication.

Linking this technical approach with analytical methods from data mining or conversation analysis, it is possible to study (using motion sensor, video and audio data) the interplay of different communicational resources (such as talk, gaze, manipulation of objects) [Dierker et al., 2009b]. Based on the empirical results, the approach allows to develop mechanisms for automated detection/anticipation of "joint attention" and shifting foci of attention that are suitable to be implemented in artificial intelligent systems and/or used the design of human-robot interaction. Moreover, using the video see-through goggles as *optional* part of the system allows to compare features of interaction under both (i) the AR-based condition and (ii) the natural condition.

In this paper, we propose a scenario that is particularly suited for interaction analysis with this system. In the first chapter, we give a short overview about software and hardware components of the system and the research questions that can be addressed with such a system. The scenario and the procedure for a first study will be described in the next two chapters and the paper closes with a short discussion and conclusion.

2 The AR-based interception and manipulation interface – Description

The AR-enabled interception interface (ARbInI) consists of two identical setups that are worn by two participants. The core of one such setup is a video-see-through head-mounted display with a front-mounted camera. Additionally, the

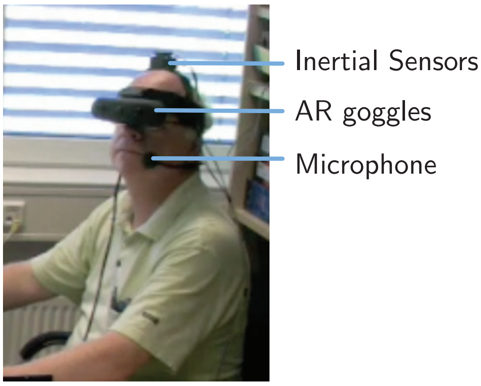


Figure 1: Equipment of one participant.

participants wear an inertial sensor on top of their heads and microphone headsets[1] (see Figure 1).

Several software components provide a solid and easily extensible basis for AR-based cooperation (see [Mertes, 2008]). Shortly sketched, the system captures the input video stream and augments virtual objects on top of physical objects using the ARToolKit marker system [Kato and Billinghurst, 1999] amongst others. In the present case, the virtual objects are wedge-shaped 3D objects with pictures on both lateral surfaces. Displaying the output video stream on the head-mounted displays closes the interaction loop. Meanwhile, all sensor data and system information is published directly over the network via the XML-enabled Communication Framework (XCF, [Wrede et al., 2004a]). An Active Memory Interface [Wrede et al., 2004b] ensures the storage of all recorded data in a shared dataset that can be analysed online and offline [Dierker et al., 2009a]. The system optionally applies machine learning methods, namely Ordered Means Models [Großekathöfer and Lingner, 2005], for the automatic recognition of head motions. This can be used to analyse and annotate head motion data according to the four motion classes: "shake", "nod", "tilt" (and the less often used "look left/right") [Wöhler et al., 2010]. The sound data from the microphones is analysed with a speech recognition software to obtain speech times for a turn taking analysis [Fink, 1999].

As another optional feature of the system, we developed an artificial communication channel to mediate attention: A multimodal gaze direction display. We control the colour of the virtual objects, which are augmented into the scene according to their position in the partner's field of view. For example, objects in the centre of participant A's field of view are coloured red for participant B while the objects in participant A's peripheral view are coloured yellow for participant

[1] The system integrates more sensors than this (e.g. headphones, wii Remotes, Vicon, underdesk camera) but these have not been used for this work.

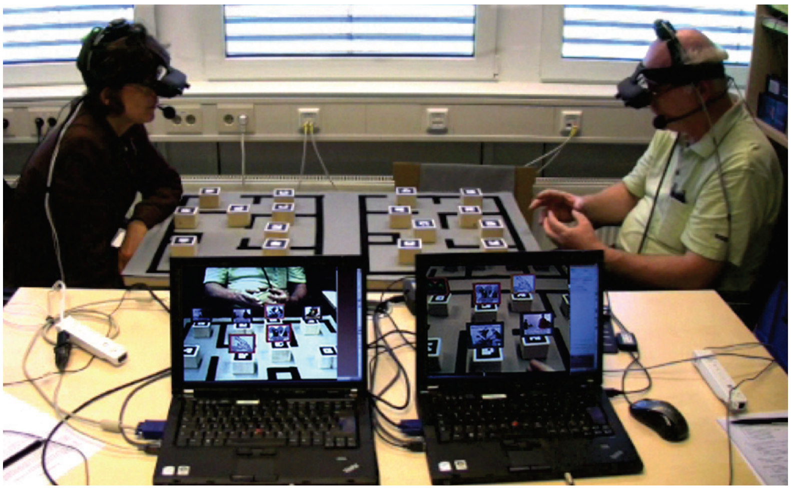


Figure 2: Example of a discussion at the beginning of the collaborative phase.

B. Objects that are not in the field of view of A are grey for participant B (see Figure 3a). Conversely, participant A sees visual augmentations of B's field of view [Mertes et al., 2009]. Since we were able to show an improvement for reaction times and error rates and received positive feedback from the participants [Dierker et al., 2009b], we provided the attention augmentation also in the present scenario in order to test its usefulness in a scenario which guides the participants' focus not on this particular feature but on other parts of the scenario.

In order to further document the user's behaviour during a study, we use a set of five scene cameras that capture the participants from different perspectives: one camera is located above the desk, monitoring the placement of the objects on the table, two cameras capture each a frontal view on one of the participants, one camera captures the video streams of both HMDs and the last camera captures the whole scene from a lateral angle.

3 Research Methodology / Usage

The technical infrastructure described above allows the investigation of several phenomena that are described in the following listing (see also Hermann and Pitsch [2009]):

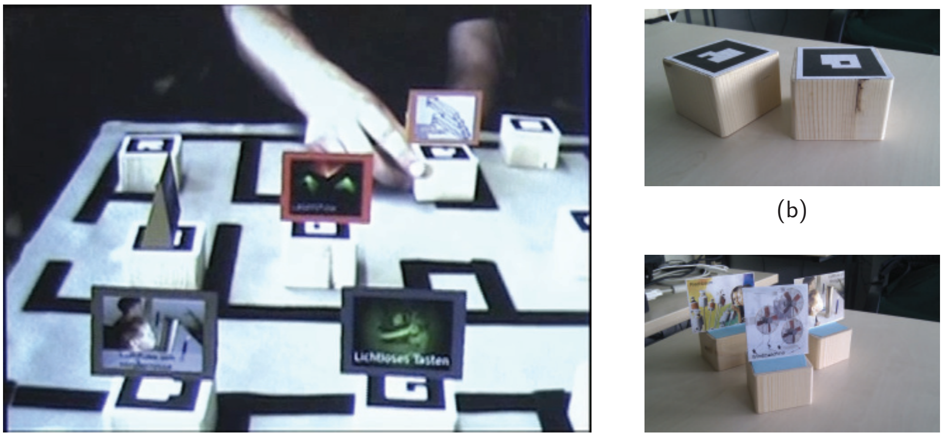
Every setup component can be omitted in order to investigate its influence on the interaction: By comparing the users' behaviour when using the setup with and without the AR goggles, we can examine the way in which the AR technique itself influences the participants' behaviour. For example, several participants complained about weight of the head-mounted displays, the noticeable lag in video display, and a narrow field of view. These differences might affect the way the users move their heads: The weight together with the lag might restrict the head movements since the users possibly want to steady the video stream. On the other hand, a compensation strategy for the decreased field of view might be increased head movements while searching for objects on a table, for example.

Secondly, the information from one user can be provided to the other user. For example, as we described above, the colour of the virtual objects' border for one user changes according to the objects' locations in the field of view of the other user and vice versa. With this, we can investigate how participants deal with information presented in a new and unknown way, if they are able to use that information effectively, and how they get accustomed to it.

Thirdly, it is possible to analyse the interaction between the participants, using their head movements, speech times or focus of attention. The system is particularly suited to investigate the mutual timing of these behaviours and their contribution to turn-taking. For example, we can quantitatively investigate the interplay of verbal backchannelling and head gestures for the turn-taking mechanisms.

Finally, the use of AR allows to manipulate the perceptible information. The simplest example is to modify the virtual information provided by the system (the visual appearance of the objects that are augmented into the scene or the timing or characteristics of the displayed attention focus of the partner). For example, the objects in the shared interaction space could be different for the two participants causing misunderstandings in the interaction. Moreover, since the system is intercepting the information flow for both the visual and the auditory domain, it is also possible to change the characteristics of the participants' speech. The system allows the timing, tone, or understandability of the partner's speech to be altered before providing it to the user. This investigates how interaction partners cope with misunderstandings and disturbances.

In sum, the ARbInI enables us to empirically investigate the impact of human multimodal conduct (e.g. head gestures, visual attention etc.) by systematically manipulating in real-time their characteristics (e.g. objects presented) and also their precise timing and frequency as controlled variables in interaction studies. However, by using AR goggles, the user's upper face is covered, leaving gaze and parts of the user's facial expressions (e.g. emotions) unavailable as communicational resources to the interaction partner. Consequently, the current system is limited in its usability for examining interaction in which participants generally face each other (e.g. small-talk). Instead, the system is particularly suited for task-oriented cooperation settings where the attention of both interactants is



(a)

(b)

(c)

Figure 3: Presentation of the stimuli: (a) virtual objects are displayed on top of cubes. The objects' borders are coloured according to their position in the partner's field of view: red border: middle of the partner's field of view, yellow/orange/green border: outer field of view, grey border: object not in the partner's field of view. (b) wooden cubes with ARToolKit markers without virtual objects. (c) for the non-HMD condition the pictures are printed on cards that are affixed to wooden cubes.

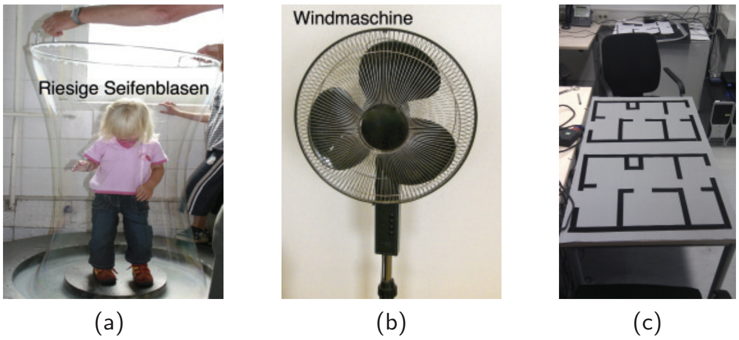
mostly oriented towards a shared space, for example, a set of objects on the table (Gaver et al. [1993], Billinghurst and Kato [1999]).

4 The Scenario

In order to compensate for the specific conditions of the AR technique (i.e. participants wearing head-mounted displays, see Fig.1), we here propose to use a task-oriented setting, in which two users have to manipulate a range of objects while talking to each other. They are asked to negotiate a solution for their problem: to design a museum exhibition (Mondada [2006], Pitsch and Krafft [2010], Heath et al. [2009]).

In this scenario, two participants are sitting face to face at a table. A floor-plan of a museum with different rooms for an exhibition is placed on the table.

The participants are asked to plan an interactive exhibition. There are 16 pictures of the exhibits that are to be distributed on the floor-plans in two tasks. In the completed museum, the exhibits would interactively teach physical characteristics of the world to the visitors. For example, one picture shows a child inside a huge



(a) (b) (c)

Figure 4: Examples of exploratory exhibits for the scenario with titles in German (a)The picture with the (translated) title *huge soap bubbles*. (b) The title is *wind engine*. (c) The floor-plans used in the Scenario.

soap bubble representing an experiment to learn about surface tension. Another experiment allows the visitor to learn about airflow by means of a huge fan (Figure 4). All 16 pictures of the experiments are labeled in German. Table 1 gives the complete list of the pictures of experiments used for the scenario [2].

The pictures/exhibits are to be placed on the floor-plan in such a way that they do not interfere with each other (see Figure 2 for a finish state for one participant pair). Most of the pictured exhibits expect certain facts for their surroundings. For example, the experiment with the huge soap bubbles needs a room where wind is unlikely. Likewise, the huge fan might be damageable by water. Thus, it should be placed in a room with no water. On the same time, the experiments also can be the source of interference. For our example this means that the huge fan emits wind and thus should not be located next to the soap bubbles experiment to avoid the destruction of soap bubbles. Moreover, visitors might play with the water of the soap bubbles experiment and thus, electric power (like it is used by the huge fan) should be nowhere in the near. Pre-tests and discussions revealed the requirements that are listed in Table 1 as well as their emissions.

5 Procedure

This chapter describes a procedure for this scenario, recently used in a study. After describing the scenario we discuss possible research questions that can be answered using this scenario and procedure.

[2] The used exhibits have been derived with permission from Phänomenta Peenemünde `http://www.phaenomenta-peenemuende.de`, Phänomania Essen `http://www.phaenomania.de/essen` and Phänomenta Lüdenscheid `http://www.phaenomenta.de/Luedenscheid`)

Title of the experiment	requires	emits
Color Mixtures	controlled lighting	Light
Extinguish candle by drumbeat	room without wind	Sound, Smell
Feel around in the dark	absolute darkness	
Humming stone	silence	
Huge soap bubbles	room without wind	Water
Lasershow	darkness	Light
Listening	silence	
Optical illusion: Swivel disk	lighting	
Optical illusion: Triangle in House	lighting	
Optical illusion: Arrows	lighting	
Plasmadisk - electric discharge		
Smelling tree	neutral smell	
Steadfast candle	room without wind	Smell/Fume
Soundfigures of sand	room without wind, dryness	Sound
Water-sound dabbling bowl	lighting	Sound, Water
Wind engine	dryness	Wind

Table 1: Titles of experiments that have been used in the scenario (the given title is a translation). The other columns list the experiments' requirements and possible sources of interference between the experiments.

In order to learn more about the influences of AR on the behaviour, we compared an AR condition (the subjects wore an HMD plus the sensors) with a non-AR condition (subjects wore only the sensors). The procedure consists of three parts and a questionnaire: after the participants are equipped with the sensors according to their respective condition, the *familiarisation phase* begins where the participants get used to the video recording and the wearable devices and meanwhile get to know each other. After five minutes, the experimenter asks the participants if they feel now comfortable and familiar with the sensors (and the HMD) or if they need more time for the acclimatization. If the participants agree to continue, they proceed with the *individual phase*. The experimenter places a floor-plan in front of each participant. Both floor-plans are identical and oriented in the same way as it is shown in Figure 4c. In order to ensure the correct placement on the table, the floor-plan is glued with black textile tape on grey cloth while the cloth is attached to the table using hook-and-loop tape. The 16 exhibits described in the scenario chapter are divided into two predetermined sets of 8 objects. Each participant is asked to unhurriedly find an arrangement for his or her 8 exhibits alone. The participants can not see their interaction partner and his/her floor-plan during this part of the trial because of a barrier in the middle of the table. Once the participants state to be finished with this task, the experimenter removes the barrier so that the participants can see each other again. In a third part, the *dyadic* *phase*, the participants are asked to discuss their arrangement of the exhibits with their partner and find, again unhurriedly, a joint solution for all 16 exhibits in one of the floor-plans. The experimenter additionally explains a helpful feature of the ARbIMI: the participants are offered a visual highlighting of the view field of their interaction partner. Afterwards, the participants are asked to complete a *questionnaire*.

Sample For this study, we recruited people that subscripted themselves to a list of potential participants before. The participants were asked to choose an appointment from a doodle poll. Each appointment needed two subscriptions. Thus, the pairs of collaborating participants were created by the participants themselves.

We tested 13 pairs of participants whose age ranged from 18 to 75 years. 11 participants stated to be computer scientists or students of computer science. 7 pairs attended the AR condition, their mean age was 30,03 years and 5 of them were female. The remaining 6 pairs attended the non-AR condition, their mean age was 30,33 years and 4 of them were female. Unfortunately, we had to leave out the data from one participant pair of the AR condition because one of the participants chose to stop wearing the HMD after the familiarisation phase because of a sudden feeling of nausea. The remaining 23 participants were analyzed.

The participants needed between 20 and 40 minutes to complete the three tasks without the questionnaire (which was not included in the time measurements). The average time was 31 minutes.

Research Questions For this study, we used the scenario described in the prior chapter in two different tasks: In the individual phase both participants worked with their own copy of the museum plan with only 8 of the 16 objects while in the subsequent dyadic phase all 16 objects are to be placed in only one plan. This way, the participants get to know a subset of the objects and their requirements very closely before they start the discussion about all 16 objects with their partner. From preliminary tests we got the impression that this acquired knowledge allows for a deeper discussion in the dyadic phase. The two floor-plans are oriented in the same way in order to facilitate the dyadic phase. Otherwise, the participants would be forced to rotate the plan in their mind. Moreover, these two different tasks together with the familiarisation phase in which the participants are free to talk about a topic they like, allow us to compare the participants' behaviour in three different interaction scenarios: A small-talk situation, a situation where the user is focussed on the table solving a problem and getting to know the exhibits and finally a collaborative task that includes exchange of information, possibly discussion as well as work with objects in a shared space (move objects on the table, hand over objects).

Using AR, we gain several important advantages: Since all objects are virtual objects displayed on top of markers, we can easily track all objects in space and over time using the head-mounted cameras. This allows us to record the position of the objects in the field of view of the users as well as who handles which object at which moment. The displayed augmented objects can be highlighted according to their position in the partner's field of view and we can easily exchange the objects even during an ongoing experiment.

Moreover, we can add meta-information next to the objects in the users' visual field. For example it is possible to display the noise level of a room resulting from the objects that are already positioned in it. This information can be displayed using icons or text. Apart from that, we can also use a multimodal display for the noise level of a room: Using additional headphones, we can allow the users to hear the soundscape at any location on the plan by interactively providing the acoustic contributions that the exhibits make.

Finally, we will be able to control which information will be perceivable by which participant. For example, we can present different features of the object to the two participants (small vs. big, silent vs. noisy, etc.), so that we are able to induce potentially problematic situations which will allow us to investigate how participants deal with such non-obvious misinterpretation of the setting.

6 Conclusion

In this work, we proposed a scenario for analysis of interaction mediated by AR. The scenario is used with an augmented-reality-enabled Interface (ARbInI) that allows us to intercept (monitoring and recording) and manipulate (disturb, enhance) the interaction using the system. More precisely, the ARbInI enables us to control which information will be perceived by which participant. Moreover, presenting different features of the object to the two participants (small vs. big, silent vs. noisy, etc.), so that we are able to induce potentially problematic situations which will allow us to investigate how participants deal with such non-obvious misinterpretation of the setting. Using the proposed scenario with this system (a) we can easily track all objects in space and over time and record who handles each at which moment, (b) we can easily adjust the displayed augmented objects, (c) we can add meta-information next to the objects in the users' visual field, and (d) we can explore truly multimodal interactions, such as allowing users to perceive the soundscape at any location on the plan by interactively mixing the

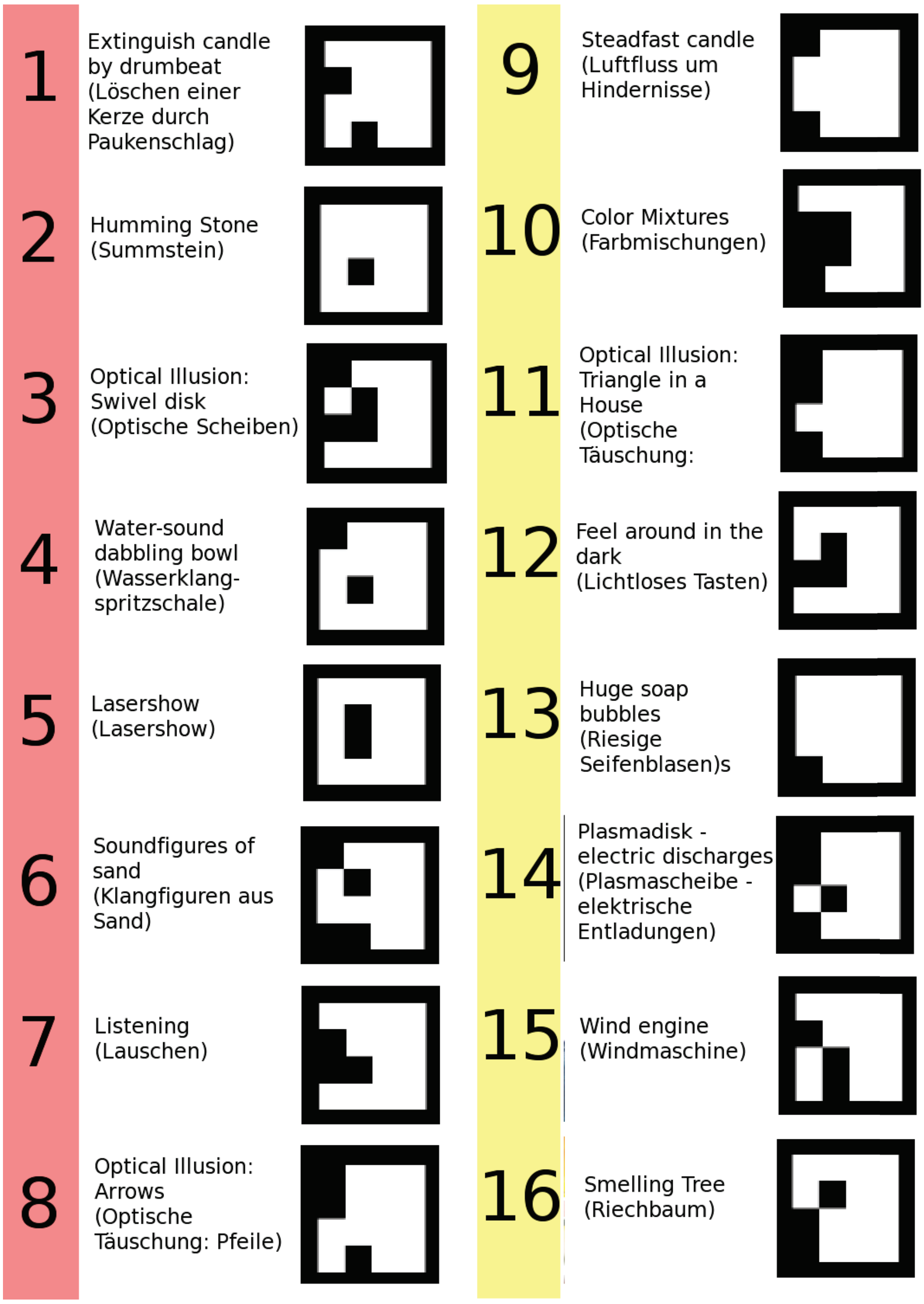

1 Extinguish candle by drumbeat (Löschen einer Kerze durch Paukenschlag)
2 Humming Stone (Summstein)
3 Optical Illusion: Swivel disk (Optische Scheiben)
4 Water-sound dabbling bowl (Wasserklang-spritzschale)
5 Lasershow (Lasershow)
6 Soundfigures of sand (Klangfiguren aus Sand)
7 Listening (Lauschen)
8 Optical Illusion: Arrows (Optische Täuschung: Pfeile)
9 Steadfast candle (Luftfluss um Hindernisse)
10 Color Mixtures (Farbmischungen)
11 Optical Illusion: Triangle in a House (Optische Täuschung:
12 Feel around in the dark (Lichtloses Tasten)
13 Huge soap bubbles (Riesige Seifenblasen)s
14 Plasmadisk - electric discharges (Plasmascheibe - elektrische Entladungen)
15 Wind engine (Windmaschine)
16 Smelling Tree (Riechbaum)

Figure 5: Mapping of displayed experiments to a set ARToolKit Markers. The used exhibits are derived with permission from Phänomenta Peenemünde `http://www.phaenomenta-peenemuende.de`, Phänomania Essen `http://www.phaenomania.de/essen` and Phänomenta Lüdenscheid `http://www.phaenomenta.de/Luedenscheid`.

acoustic contributions that the exhibits make. Finally, we described a study using the scenario.

Acknowledgements

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).)

Bibliography

M. Billinghurst and H. Kato. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, pages 261–284. Citeseer, 1999.

Angelika Dierker, Till Bovermann, Marc Hanheide, Thomas Hermann, and Gerhard Sagerer. A multimodal augmented reality system for alignment research. In *International Conference on Human-Computer Interaction*, pages 422–426, San Diego, USA, 18/07/2009 2009a.

Angelika Dierker, Christian Mertes, Thomas Hermann, Marc Hanheide, and Gerhard Sagerer. Mediated attention with multimodal augmented reality. In *ICMI-MLMI '09: Proceedings of the 2009 international conference on Multimodal interfaces*, pages 245–252, New York, NY, USA, November 2009b. ACM. ISBN 978-1-60558-772-1. doi: http://doi.acm.org/10.1145/1647314.1647368.

Gernot A. Fink. Developing HMM-Based Recognizers with ESMERALDA. In *TSD '99: Proceedings of the Second International Workshop on Text, Speech and Dialogue*, pages 229–234, London, UK, 1999. Springer-Verlag. ISBN 3-540-66494-7. URL http://portal.acm.org/citation.cfm?id=647237.720414.

W.W. Gaver, A. Sellen, C. Heath, and P. Luff. One is not enough: Multiple views in a media space. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 335–341. ACM, 1993. ISBN 0897915755.

Ulf Großekathöfer and Thomas Lingner. Neue Ansätze zum maschinellen Lernen von Alignments. Master's thesis, Bielefeld University, September 2005.

Christian Heath, Paul Luff, and Karola Pitsch. *Indefinite precision: the use of artefacts-in-interaction in design work*, pages 213–224. Routledge Chapman, London, 2009.

Thomas Hermann and Karola Pitsch. C5: Alignment in AR-based cooperation. In *Funding Proposal for the 2nd period of the Collaborative Research Centre SFB 673 "Alignment in Communication"*, pages 357–378. Bielefeld University, 2009.

H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, volume 99, pages 85–94. San Francisco, CA, 1999.

Christian Mertes. Multimodal augmented reality to enhance human communication. Master's thesis, Bielefeld University, 2008. URL `http://bieson.ub.uni-bielefeld.de/volltexte/2009/1414/`.

Christian Mertes, Angelika Dierker, Thomas Hermann, Marc Hanheide, and Gerhard Sagerer. Enhancing human cooperation with multimodal augmented reality. In *International Conference on Human-Computer Interaction*, pages 447–451, San Diego, USA, 18/07/2009 2009. Springer.

L. Mondada. Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse studies*, 8(1):117, 2006. ISSN 1461-4456.

G. Papagiannakis, G. Singh, and N. Magnenat-Thalmann. A survey of mobile and wireless technologies for augmented reality systems. *Computer Animation and Virtual Worlds*, 19(1):3–22, 2008. ISSN 1546-427X.

Karola Pitsch and Ulrich Krafft. *Von der emergenten Erfindung zu konventionalisiert darstellbarem Wissen. Zur Herstellung visueller Vorstellungen bei Museums-Designern*, pages 189–222. de Gruyter, Berlin, 2010.

I.E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

C. Ulbricht and D. Schmalstieg. Tangible augmented reality for computer games. In *Proceedings of the Third IASTED International Conference on Visualization, Imaging and Image Processing*, pages 950–954, 2003.

Nils-Christian Wöhler, Ulf Großekathöfer, Angelika Dierker, Marc Hanheide, Stefan Kopp, and Thomas Hermann. A calibration-free head gesture recognition system with online capability. In *International Conference on Pattern Recognition*, Istanbul, Turkey, 23/08/2010 2010.

S. Wrede, J. Fritsch, C. Bauckhage, and G. Sagerer. An xml based framework for cognitive vision architectures. In *Proc. Int. Conf. on Pattern Recognition*, volume 1, pages 757–760, 2004a.

S. Wrede, M. Hanheide, C. Bauckhage, and G. Sagerer. An active memory as a model for information fusion. In *Proc. 7th Int. Conf. on Information Fusion*, volume 1, pages 198–205, 2004b. URL `citeseer.ist.psu.edu/wrede04active.html`.

S. Wrede, M. Hanheide, S. Wachsmuth, and G. Sagerer. Integration and coordination in a cognitive vision system. In *Int. Conf. on Computer Vision Systems*, 2006.