S. Bødker, N. O. Bouvin, W. Lutters ,V. Wulf and L. Ciolfi (eds.)
ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work,
24-28 September 2011, Aarhus, Denmark
© Springer 2011

Collaboration in Augmented Reality: How to establish coordination and joint attention?

Christian Schnier, Karola Pitsch, Angelika Dierker, Thomas Hermann
Faculty of Technology, Applied Informatics, Bielefeld University, Germany
{cschnier}{kpitsch}{adierker}{thermann}@techfak.uni-bielefeld.de

Abstract. We present an initial investigation from a semi-experimental setting, in which an HMD-based AR-system has been used for real-time collaboration in a task-oriented scenario (design of a museum exhibition). Analysis points out the specific conditions of interacting in an AR environment and focuses on one particular practical problem for the participants in coordinating their interaction: how to establish joint attention towards the same object or referent. Analysis allows insights into how the pair of users begins to familarize with the environment, the limitations and opportunities of the setting and how they establish new routines for e.g. solving the 'joint attention'-problem.

Introduction

Over the last 15 years a range of initiatives has emerged that develop and explore Augmented Reality (AR) systems, in which the user's perception of the world is overlayed with additional, digital information (Caudell & Mizell 1992; Azuma 1997). Most commonly, these systems focus on augmenting the user's visual perception by video taping in real-time the user's environment and displaying this image together with overlayed additional information on a screen. To achieve this effect, existing AR-systems either (i) exploit the cameras/displays of recent mobile phone technlogies or (ii) equip the user with specialized glasses, so-called headmounted displays (HMD). The first approach benefits from using an already available technology and easy integration into the user's everyday practices, which is reflected in the current boom of applications for navigation, interactive tourist guides etc. The second approach allows to support richer and more complex activities, during which the user could freely use his hands to manipulate objects, which is relevant e.g. in aircraft maintainance where 3D construction plans are made available, in situ, to the engineer. Whilst existing research predominantly focuses on individual users, little is known about AR-technologies in collaborative settings.

If we want to explore AR-systems using HMDs supporting real-time collaboration of physically co-present interaction partners, this creates particular conditions for the interaction: In comparison to natural face-to-face interaction, to wear HMDs and see the world through its lenses, results in limited access to usually available communicational resources: reduced field of view (Arthur 2000), lower resolution (Azuma 1997) and problems in determining the co-participant's focus of attention (Brennan et al. 2008). Additionally, looking through HMDs results in significantly less eye rotation and increased head orientation when attempting to focus on a given point or object (Kollenberg et al. 2008). It seems that a new prototype of mediated co-present face-to-face interaction emerges that places particular demands on the ways in which users can collaborate and organize their joint actions. A range of empirical questions and technical challenges arises: How can co-participants, under these conditions, organize their interaction and coordinate their activities? How can they establish joint attention? How could we design such collaborative AR-systems in a way as to substitute for the technological constraints and support the users' collaboration?

In this paper, we will present some initial findings from a quasi-naturalistic AR-experiment, in which we have equipped pairs of users with HMDs and asked them to jointly design a museum exhibition while arranging a set of objects (the exhibits) on a given floor plan. Our analysis will address the questions raised above and – using sequential micro-analysis stemming from Conversation Analysis – focus (1) on the specific interactional conditions that arise from this setting, (2) the ways in which users organize their (inter-)action and (3) how they start to establish new collaborative orientation routines. Finally, we will discuss the analytic results with regard to insights into communicational procedures and implications for the design of collaborative AR-systems.

Background

Endeavours to support realtime remote collaboration in the workplace have seen the development of a wide range of novel technologies, such as video-conferencing systems, media spaces or collaborative virtual environments. Empirical investigation of such systems has revealed the limitations of such technologies in comparison to unmediated face-to-face interaction: "gaze, gesture and other body movements are generally not as effective as in normal face-to-face communication" (Yamashita et al. 1999). When designing such systems, particular challenges consist in dealing with time delays caused by the technical transmission and the interdependencies of action and the physical environment. Participants in social interaction orient themselves and others in the local environment, refer to objects and its specific features, and attempt to animate and transform these for their practical purposes at hand – aspects which have shown to be highly problematic in technically mediated interaction: "the system fractures the environments of action and inadvertently undermines the participants' ability to produce, interpret, and coordinate their actions in collaboration with each other" (Luff et al. 2003: 53). For a collaborative virtual environment – which shows a range of parallels with the setup used in this paper – Fraser et al. (2000) show that differences between the experience of virtual environment and physical reality occur, e.g. reduced field of view, no haptic feedback, and technical network delays in VR setups. They suggest to render the limitations of the technology visible to the user, e.g. by artificially providing – similar to our system – information about the co-participant's field of view via an additional object projection or by giving haptic feedback in other media. Thus, comparison between the ways in which users deal with such augmentations in AR- vs. VR-setups will be interesting to investigate (Milgram & Kishino 1994).

While a few collaborative AR-setups have been proposed (Schmalstieg et al. 1996, Billinghurst et al. 2002, Dierker et al. 2009), little is known yet as to how participants can deal with the conditions implied by the technical constraints when attempting to fulfill a joint task. In fact, with the collaborative HMD-based AR-scenario, a new prototype of face-to-face communication seems to arise: On the one hand it encompasses aspects typical of face-to-face interaction: physical co-presence, shared interaction space, participants can touch, smell and hear each other and they can jointly manipulate the same objects. On the other hand, participants see the world through the eyes of a videocamera with a reduced field of view and – due to cost and computing power – mostly monoscopic vision, and virtual augmentations are not necessarily similar for both co-participants; these features are comparable to technologically mediated settings. In such co-present, but technologically mediated setting, the co-participants are faced with the task of organization their multimodal interaction (Goodwin 2000): to coordinate their actions (Deppermann & Schmidt 2006), to monitor and take into account the co-participant's current state of action (Goodwin 1980) and to establish joint attention (Kaplan & Hafner 2006). From this, the empirical questions arise as to how users can interact with each other under these specific conditions, and to which extent they might adapt or develop new procedures and interactional strategies.

AR-System for collaborative task-oriented interaction

Over the last years, we have developed an AR-system, that allows for real-time collaboration of two users and to record, intercept and manipulate the users' natural communication channels. The system encompasses the following components: HMDs with an integrated camera that captures the view from the user's perspective and passes it on as a video frame that is projected on the screen of the corresponding HMD. Similarly, audio signals can be captured with microphones and relayed via in-ear headphones. This paradigm allows to precisely record the relevant sensory information available to interacting users. This enables us to reconstruct the user's audio-visual perceptions and to gain a better understanding of their respective member's perspective in co-present interaction. We furthermore record the detailed head movements by inertial sensors worn on the head. This allows us to measure accurately amplitude, timing of head gestures such as nodding and head shaking.

Secondly, we can manipulate (modify or augment) the information streams and thus study effects of disturbances, ranging from enduced color-blindness to completely different scenes the users perceive when looking at one at the same physical object. Beyond this basic *Interception & Manipulation* functionality, we augment virtual objects on top of physical objects in the interaction space using ARToolkit (Kato & Billinghurst 1999). Specifically, for the museum planning scenario, these virtual objects are exhibit pictures shown on wedge-shaped 3D objects. As novel contribution we introduced a coupling of users by a joint-attention-support channel: each user sees by coloring of the exhibit objects whether and how much they are in the view field of their interaction partner. More precisely, the object's frame color changes from yellow (peripheral) to red (in the center of the partner's field of view). Details on this augmentation are explained in Dierker et al. (2009).

Experiment: Collaborative museum exhibition design

To investigate collaboration under the specific conditions of AR-technology, an experiment has been conducted (08/2010), in which pairs of users were asked to jointly plan a museum exhibition while arranging a set of objects (the exhibits) on a given floor plan (Pitsch & Krafft 2010, Luff et al. 2009, Dierker et al. 2011). The participants were seated face-to-face at a table, equipped with AR-glasses, microphone headsets and an inertial sensor on top of their heads. The participants were asked to carry out three subsequent tasks: (1) In a *familiarization phase* (5 minutes) the participants were asked to chat with their partner about a self-chosen topic in order to familiarize with being videotaped and wearing the devices. (2) In the following *individual phase* (with vision obstructing barrier) they were asked to individually plan a museum exhibition using a set of 8 different exhibits each (wooden blocks as material 'handles' for augmented objects sitting on top of the blocks), and arrange them on a given floor plan. (3) In the subsequent *dyadic phase*, the participants were asked to discuss their arrangement of the exhibits with their partner (without barrier) and to develop a joint solution for all 16 exhibits in one of the two identical floor plans.

The specific conditions of interacting in an AR-setting

The specific conditions of an AR-system place particular demands on collaborative action. Analysis will enable us to gain a first understanding of the extent to which the participants have to deal with a new situation, and, at the same time, build the basis for subsequent analysis in the following parts.

(1) *Dual ecology and the world's instability*: AR-settings present a dual ecology for the participants: On the one hand, they have physical access to the real world such as the table with wooden blocks and floor plan, and are in physical co-presence with the interaction partner. On the other hand, on the level of visual perception, this world is mediated through the display of the real-time video-stream and the added virtual augmentations.[1] During the interaction, this dual ecology is a mixture of stability and instability, e.g. the wooden block actually forms a stable reference point, but this is not the case in the augmented content. Consider now the short fragment F1 (Figure 1) showing the respective participants' view.

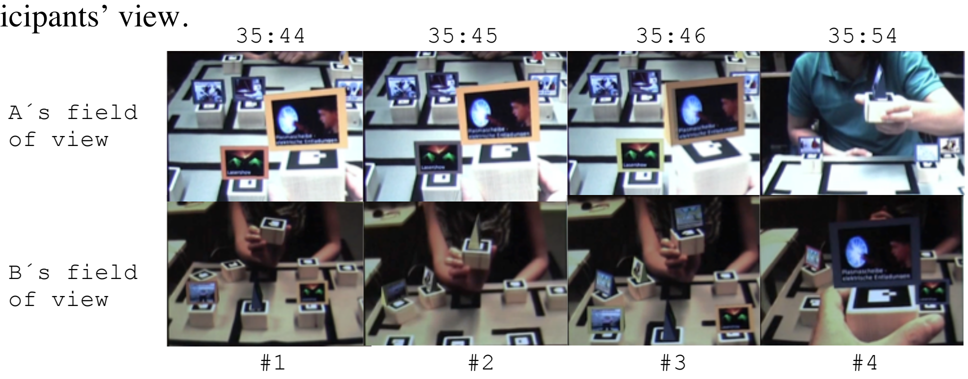


Figure 1. Participants' view

When A lifts the wooden block and presents it to B, it is visually augmented as the exhibit 'Plasmascheibe (plasma dial)' (fig. 1, #1). For participant B, however, it appears firstly as having no augmenteation, then secondly as being positioned sideways (fig. 1, #2) and thirdly, it appears as the exhibit 'Dreieck im Hause (triangles in the house)' (fig. 1, #3). Thus, at this moment of the interaction, both

[1] We borrow the term "dual ecology" from Kuzuoka et al. (2004), but re-define it. Kuzuoka et al. use the term for a remote collaboration setting to denote the ecology of the local site vs. that of the remote site.

participants are faced with *different* exhibits in the augmented world although referring to the same object in terms of the real world. Only after the object has been handed over from A to B, the object also appears as exhibit 'Plasmascheibe (plasma dial)' to him, while – now – being displayed sideways for A (fig. 1, #4). Thus, the participants cannot be sure to share the same representation of their interactional world, which will have consequences for the way in which they are able to refer to objects and establish joint attention. In this particular case, the observed instabilities are caused by irritations in the marker tracking due to obstruction and rotation of the objects as they occur in social interaction, and improvements in an iteration of the system are under way. However, in more subtle ways similar effects occur also in other AR-settings, so that participants generally have to deal with them to some extent.

(2) *(Dis-)Embodiment*: The dual ecology between real and virtual world and the lack of a stereoscopic view also influence the ways in which participants can deal with their own and the co-participant's bodily existence in the world. This becomes evident e.g. when they attempt to handle objects or exchange them with their co-participant. In the following fragment F2 (fig. 2), participant A presents an object to B, which he, in turn, attempts to grasp (cf. #1). Pic 2 and 3 show the orientation of his open, ready-to-grasp-hand, however, slightly disoriented to the side. In pic 4, he has repaired this action and orientation of the hand, so that in (#5) he can indeed grasp it (#6). Thus, the participants' interact under the condition of unusual tactile senses and awareness of the environment.

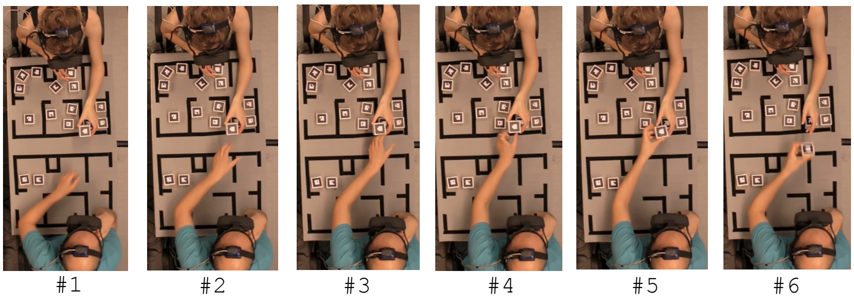


Figure 2. Grasping

(3) *Orientation and interactional coordination*: The participants have available – due to the AR-glasses – only a highly reduced field of view with a masked periphery, so that they can only either look at their co-participant or inspect (parts of) the museum plan and/or objects. Thus, focusing on the task, they are hardly aware of the partner's physical representation, body movements, head orientation etc. – aspects, which are known from face-to-face interaction, to be important resources for coordination and organizing social interaction. This places a particular demand on coordinating their actions and establishing joint attention to some object or area on the plan as shown in Fragment F3 (fig. 3). This sequence

of video-stills, i.e. #1-4, shows a point in time where participant B is looking forward to have the floor but does not have chosen a specific object group yet.

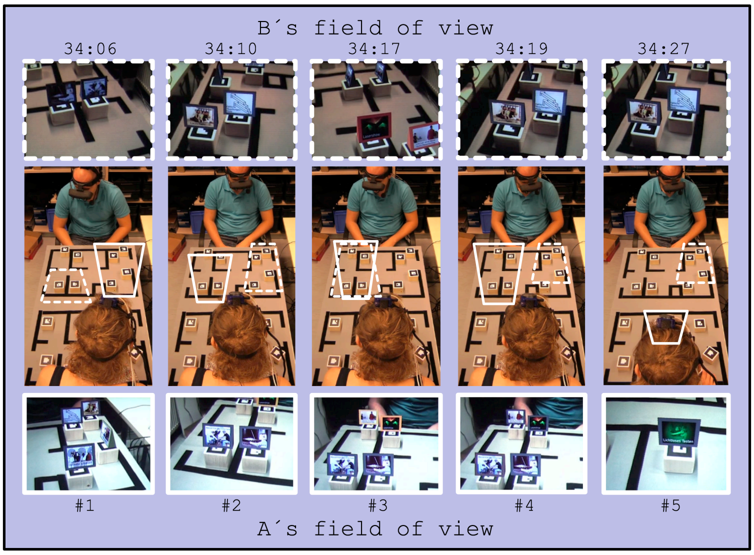


Figure 3. Participants’ difficulties in coordinating their foci of attention

However, this choice has been made in #5 but participant A is not aware of it just at that moment. Because of the impossibility to coordinate their gaze direction, both interlocutors are focused on different parts of the map, except for a random gaze overlap in image #3. The physical projections of participant B are outside of participant A’s field of view, and vice versa. Moreover, it is obvious that the technical highlighting support is not used here. The mutual orientation is indicated by the red highlighted object frame (#3). However, picture #4 and #5 show that the attention shift of subject B is not retraced by interlocuter A.

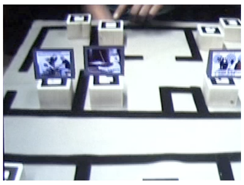
This demonstrates that the users’ head movements cannot be used as an indicator of the co-participant’s visual orientation, e.g. they are not available as communicational resources. Additionally this example suggests that orientational cues, such as leaning forward to a specific object, have only restricted interactive relevance. Similarly, the participants don’t have access to facial expressions as the HMDs hide important parts of the user’s face.

Organizing (inter-)action: Establishing joint attention

Fragment F4 occurs early in the third interaction phase (32s after beginning) and reveals the practical problems in establishing mutual orientation on some object. At the beginning participant B tries to orient his interlocuter A to the object ‘lasershow’, which he attempts to do exclusively by verbal means: “euh-well this LASERshow here;” (01). However, this verbal deictic reference turns out to not be sufficient for a precise reorientation of the co-participant. A does indeed not change her orientation, and answers with an elongated “hm:=hm”, which participant B treats as only ‘claiming’, but not ‘showing’ understanding. B correctly treats her answer as not having followed his orientation and reformulates his suggestion. He adds “you see it here,” (02). This time, he adds a gestural pointing to the object (#1) and designs his turn as a question, which projects the co-participant’s confirmation. A indeed reacts to this second attempt by leaning forward and commenting on the indicated object.

Fragment F4 (32:53-33:00)

```
01 B:        |äh=also diese LASERshow hier;| (0.2)
              euh=well this LASERshow here;

02 A:        |hm:=hm:-|
   B: -->    |und die |(0.2)|siehste hier,|(0.5)
              and the         you see it here,
                                         #1

03 B:        |in diesem RAUM hab ich den extra reingestellt,|
               I put it in this ROOM for a good reason,

04 B:        |<<p>mh=weil>|
              mh=because
   A:        |warte,      |
              please wait,
```

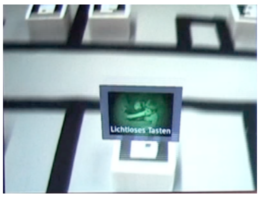

#1

A second fragment (fragment F5) allows us to investigate the procedures of orienting attention in greater detail. It follows directly on fragment F3 (fig. 3), for which we have shown that the participants are oriented to different parts of the floorplan without being aware of the other’s orientation.

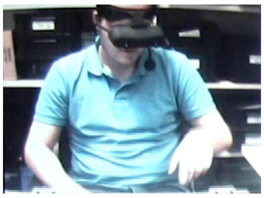
Fragment F5 (34:28- 34:35)

```
01 B:  -->   |<<all>.h |zum Beispiel HIER in der=in der> (.)
                      #1 for example HERE in the=in the

02 B:  -->    also von dir aus| OBEREN rechten ecke-|
              that is to say from your perspective in the UPPER right corner-
   A:                                                 |hm?|
                                                  #2

03 B:        |da sind äh: einmal diese pfeile-|
              there are er: once this arrows-
                        #3
```

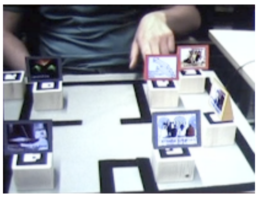
A´s field of view


#1, 34:30

A´s field of view


#2, 34:33

A´s field of view


#3, 34:34

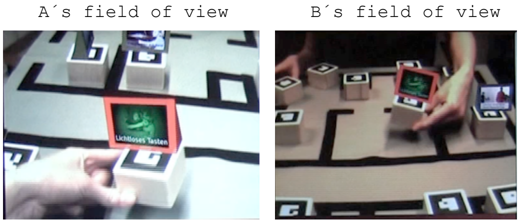
When participant B now attempts to orient B to a particular object using a combined verbal and gestural procedure – “for example HERE in the=in the” and pointing to the object – participant A is oriented to a different part of the floorplan (#1). B then repairs his action and adds a precise verbal localization: “that is to say from your perspective in the UPPER right corner" (02). It is only as a reaction to this explicit localization that A begins to re-orient. Interestingly, she looks to B's face (rarely done in our setup), sees him being oriented to his left side (#2) and then follows his gesture to the indicated location on the plan (#3). The participants thus are faced with the problem of formulating places in a way that they can be subject to a locus of shared visual focus with the co-participant. While interacting, they experience that they cannot rely on well-established procedures from their daily life, but have to be attentive to the lack of their co-participant's reaction and repair their actions to provide more information than usual. At the same time, they don't appear to repair at each stage, which means that they have to interact under the assumption that – if not otherwise signaled – their co-participant is following their actions.

Experimenting procedures and establishing routines

While the participants at the beginning of the collaborative part have to familiarize themselves with the specific conditions of the setting, during the interaction they can be seen to adapt to the specific needs and how to best collaborate with their interaction partner. In this vein, the AR-setting proposes to be a valuable setting for investigating how participants exploit the constraints of the technology and how these may lead to emerging new interactional routines. We will explore this issue by presenting the following fragment F6, in which we continue our focus on the participants' practices around establishing joint attention.

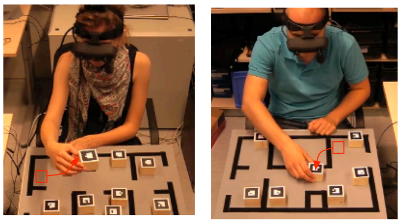
Two minutes after fragment F4, the pair of users comes to a decisive turning point in the ways in which they attempt to orient the co-participant's visual focus of attention: They give up the exclusive use of verbal deixis or verbal-plus-gestural-deixis (which have turned out to be problematic and in need of repair if no explicit point of common reference has been established before, see previous section). Instead, participant A introduces a new procedure: She lifts the object from the floor plan and presents it slightly elevated for a longer stretch of time (4 seconds), and detects that her co-participant is able to follow her orientation easily, so that – differently from the procedures shown in the previous section – no repair is required.

Fragment F6



#1, 35:02 #2, 35:02

Fragment F7



#1, 36:45 #2, 36:47

In subsequent fragments, we find her re-using this procedure several times, while the previously deployed strategies diminuish. In re-using this procedure, she experiments with different ways of perfoming it until a new orientational routine has emerged that is well-functionning 'for all practical purposes' for these participants: lifting the object with more pronounced hand movements.

Thus, the participants' first attempts of establishing a joint focus of attention have provided them with a practical experience that enables them to gain a conception of the actual possibilities and limitations imposed by the technology and the ensuing specific interactional conditions.

The following fragment F7 shows this new orientation method in a distinct form as it emerges for the first time. Participant A lifts the object out of the original context, shows it to participant B, and verbally specifies it by denominating the corresponding exhibit 'triangle in the house' (#1). This enables participant B to respond adequatly by selecting the appropriate object from his plan, which he, then, places in the emerging common perceptual space (#2).

While being highly functional in orienting the co-particpant's visual focus of attention to a given object/point of reference, it also helps to solve a further problem: In this case, the exhibit, which has been pointed out by A, did not appear at first in its augmented form in participant B's field of view. By slightly turning the wooden handle this problem – caused by marker orientation and obstruction – can easily be dealt with.

Discussion

We have presented an initial investigation from an experiment, in which we have used our HMD-based AR-system for real-time collaboration in a task-oriented scenario (collaborative design of a museum exhibition). We have pointed out the specific conditions, under which the participants interact with each other, namely: (i) dual ecology and the world's instability, (ii) (dis-)embodiment, and (iii) limited access to the co-participants conduct and thus lacking information for organizing the interaction. We have then focused on one particular practical problem for the participants in coordinating their interaction, namely how to establish joint attention towards the same object or referent. Analysis has revealed two procedures and the co-participants' reactions to it. A chronological investigation of the participants' procedures and methods has lead to first observations on how the pair of users begins to familarize with the environment, the limitations and opportunities of the setting and how they can use it to establish new routines for e.g. solving the 'joint attention'-problem.

This experiment presents an important practice-based trial for our technology under new interactional conditions. It performed well, robust and stable and constitutes a solid basis for further interaction experiments. However, minor technical aspects have come to light that need addressing, in particular the stability of marker tracking in an environment where users are likely to obstruct parts of the patterns.

From our analyses some general implications for the design of HMD-based AR-settings arise: We have to exploit novel ways for dealing with the limited availability of communicational resources. Most prominently, we have to provide support for allowing a participant to access the co-participant's visual orientation in space. In the current setting the implemented method has only rarely been used for orientation by the participants. Further investigation is required into the conditions, under which such orientational cues might become relevant and usable for the participants, and to think of other forms of augmentation.

Another novel aspect emerges: the AR-setup as a tool for investigating social interaction. Our AR setup enables us to systematically intercept and manipulate a range of interactional features in real-time (vision, audio). This opens the innovative possibility to modify both auditory and visual perceptual signals as controlled variables, e.g. by introducing delays, changing frequencies or specific characteristics. This allows us to address questions such as: Up to which differences do interacting users tolerate specific disturbances? Which multimodal features of the complex holistic phenomenon 'multimodal interaction' are relevant for which kind of interactional effect? How do interacting users attempt to repair or compensate such effects? How might they change their communicative procedures and develop new interactional routines?

With our combined approach which fuses (i) a technology-driven modification of the interactional situation, (ii) semi-natural social interaction, and (iii) technical measurement of some signal streams in real-time, we gain the possibility to link qualitative micro-analysis stemming from Conversation Analysis with quantitative approaches on a larger corpus.

Acknowledgments

This work has been supported by the Collaborative Research Centre 673 *Alignment in Communication*, funded by the German Research Foundation (DFG). We are indebted to Ulrich Dausendschön-Gay for invaluable discussions.

References

Arthur, K. (2000): *Effects of field of view on performance with head-mounted displays*. Dissertation, University of North Carolina.

Azuma, R. (1997): 'A Survey of Augmented Reality', *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, August 1997, pp. 355-385.

Billinghurst, M. and Kato, H. (2002): 'Collaborative Augmented Reality', *Communication of the ACM*, vol. 45, no. 7, 2002, pp. 64-70.

Brennan, S., Chena, X., Dickinsona, C., Neidera, M. and Zelinsky, G. (2008): 'Coordinating cognition: The costs and benefits of shared gaze during collaborative search', *Cognition,* vol. 106, no. 3, March 2008, pp. 1465-1477.

Caudell, T. and Mizell, D. (1992): *Augmented Reality: An Application of Head-Up Display Technology to Manual Manufacturing Processes*, Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. 2, January 1992, pp. 659-669.

Deppermann, A. and Schmitt, R. (2006): 'Koordination. Zur Begründung eines neues Forschungsgegenstands', in: R. Schmitt (ed.): *Koordination*. Gunter Narr Verlag Tübingen.

Dierker, A., Pitsch, K. and Hermann, T. (2011): *An augmented-reality-based scenario for the collaborative construction of an interactive museum*. Bielefeld: Bielefeld University.

Dierker, A., Bovermann, T., Hanheide, M., Hermann, T. and Sagerer, G. (2009): *A multimodal augmented reality system for alignment research*, Proceedings of the 13th International Conference on Human-Computer Interaction, pp. 422–426, San Diego, USA.

Fraser, M., Glover, T., Vaghi, I., Benford, S., Greenhalgh, C., Hindmarsh, J. and Heath, C. (2000): *Revealing the Realities of Collaborative Virtual Reality*.

Goodwin, C. (2000): 'Action and embodiment within situated human interaction', *Journal of Pragmatics*, vol. 32, no. 10, September 2000, pp. 1489- 1522.

Goodwin, Marjorie H. (1980): 'Processes of mutual monitoring implicated in the production of description sequences', *Sociological Inquiry*, vol. 50, no. 3, July 1980, pp. 303-317.

Kaplan, F. and Hafner, V. (2006): 'The challanges of joint attention', *Interaction Studies*, vol 7, no. 2, pp. 135-169.

Kato, H. and Billinghurst, M. (1999): *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, pp. 85-94.

Kollenberg, T. et al. (2008). *Visual Search in the (Un)Real World: How Head Mounted Displays Affect Eye Movements, Head Movements and Target Detection*. ACM SIGGRAPH conference proceedings.

Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y. and Heath, C. (2004): *Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability.* SIGHCI conference on Human factors in computing systems.

Luff, P. et al. (2003): 'Fractured ecologies'*, Human-Computer Interaction*, vol. 18, 1, pp. 51-84.

Luff, P., Heath, C., and Pitsch, K. (2009): 'Indefinite precision: the use of artefacts-in-interaction in design work', in C. Jewitt: *The Routledge Handbook of Multimodal Analysis*, 213-224.

Milgram, P. and Kishino, F. (1994): *A Taxonomy of Mixed Reality Virtual Displays*. IEICE Transactions on Information and Systems E77-D, 9 (September 1994), pp. 1321-1329.

Pitsch, K. & Krafft, U. (2010): 'Von der emergenten Erfindung zu konventionalisiert darstellbarem Wissen. Zur Herstellung visueller Vorstellungen bei Museums-Designern', in U. Dausendschön-Gay, Ch. Domke & S. Ohlhus (eds.): *Wissen in (Inter-)Aktion*. Berlin: de Gruyter, pp. 189-222.

Schmalstieg, D., Fuhrmann, A., Szalavari, Z. and Gervautz, M. (1996). *Studierstube – An Environment for Collaboration in Augmented Reality*, in CVE ´96 Workshop Proceedings, 19-20th September 1996, Nottingham, Great Britain.

Yamashita, J., Kuzuoka, H., Yamazaki, K., Miki, H., Yamazaki, A., Kato, H. and Suzuky, H. (1999): *Agora: Supporting Multi-participant Telecollaboration*, Proc. of HCI, pp. 543-547.