Interaction Space in Technology-Enhanced Workplace:

Methods for Usability Studies

Mirja Annikki Lievonen

2015

Royal Holloway College
University of London

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

Declaration of Authorship

I Mirja Lievonen hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Mirja Lievonen

Acknowledgements

The following persons deserve sincere thanks for their encouragement and support in the course of my PhD research:

Professor John Sillince initially brought to my attention Professor Duska Rosenberg, whose research project at Royal Holloway University of London caught my interest to the point that I contacted her. The following year, I became part-time PhD student in her supervision.

I am particularly grateful to Professor Duska Rosenberg, my supervisor: despite unexpected hurdles that delayed this work, Duska was an unwavering support, critic and guide throughout these years. It was both enjoyable and challenging to link our two different perspectives onto my topic. As a result, I learned a lot.

Ralf Doerner, Professor in Computer Graphics and Virtual Reality at RheinMain University of Applied Sciences, guided me beyond my comfort zone in the project 'Interaction Spaces in the Virtual IT-Workplace' (IS-VIT). The time spent in the visualisation laboratory in Wiesbaden, Germany, was a great learning experience, thanks particularly to Ralf's constructive guidance and expertise. I shared there an inspiring time also with Guido Kuehn and Swen Walkowski: vielen Dank, Guido und Swen.

My thanks go to Dr Dave Rayner, the Project Coordinator in the 'SWan Video-conferencing usability project', to David House, the Head of AV and other members of staff, in particular the teachers who allowed me to observe their lecturing.

Professor (IIAS) Tarkko Oksala and Jouko Seppänen, the former Chief Systems Scientist at Helsinki University of Technology, gave me a lot of intellectual stimulus due to their extensive knowledge background in the human thought and history of science.

Conversations with Dr Pedro Contreras, my teammates and fellow PhD students enriched my views.

A grant from the Academy of Finland (no 209849), Finnish Cultural Foundation, and Emil Aaltonen Foundation, and travel awards from the Finnish Association of Architects, The Arts Council of Finland, and the School of Management Royal Holloway University of London, supported this work.

Professor Stanley Peters from Stanford University, Professors Trevor Harris and Maija Rautamäki from Aalto University, Dr Eeva Aarrevaara from Lahti University of Applied Sciences, Dr G. Harindranath, Dr Catherine Liston-Heyes and Dr José-Rodrigo Córdoba-Pachón from Royal Holloway wrote references to my grant applications.

I would like to express my sincere thanks also to members of staff at Royal Holloway, in particular to Jacqueline Brackenbury and David Garrod. Similarly, Chris Hitchins, Julia Charlton, Ben Gredland, Steve Henry, and Nicolas Schroeder with his colleagues at IT Help Desk provided much appreciated help. Great thanks go also to Dr Vladimir Nikiforov, Revd Sally Rogers and their colleagues for conversations, and to the choral and organ scholars for the performances that contributed to my wellbeing at Royal Holloway.

I am particularly indebted to my friends Hari and his family, Diego, Kirsti, Pirkko, Pirjo and Maini for sharing my joys and burdens over these years.

I remember my parents with deep gratitude for their love, care and work. The encouragement of Olavi and Leena, my brother and my other sister, who died while this thesis was in progress, meant a lot to me. So did also Maila's encouragement. I dedicate the thesis to Jukka and Outi, my nephew and my niece.

Abstract

The thesis brings together Architecture, Cognitive Psychology and Human-Computer Interaction to investigate the role of space in collaboration and to inform workplace design in interaction spaces. Motivation arises from technology-enhanced collaboration that poses new issues for spatial design.

Modal integration is a key concept to discriminate between collocated and ICT-mediated collaboration: People effortlessly locate/identify objects of deictic reference in a collocated situation. In a video-mediated situation, the triangle between the speaker, the addressee and the object of reference becomes fragmented. Therefore, it is difficult to infer the location of an object indicated by a remote participant and to establish joint attention. While shared understanding is necessary for collaboration, the challenge for workplace design is to design cognitively ergonomic interaction spaces.

Deixis is identified in the thesis as a key concept in the investigation of human communication /collaboration /navigation, because it bridges the cognitive actor with the object and subjective experience with the physical setting. When situated, it locates modal integration in the position of an actor, and it connects participants' perspectives on a common object. Thereby, it becomes a tool in the design of technology-enhanced work settings.

In the framework proposed, *deixis* is situated in the communicative situation by defining the position of the participant through the human-environment interface (HEI), and the interpersonal connection by the human-human interface (HHI). Situated deixis renders the interaction layout in terms of spatial, social, modal and cognitive dimensions. Therefore, it guides the coordination of complementary perspectives (e.g. speaker / addressee) in the design of hybrid work settings.

The framework was demonstrated feasible in the investigation of gestural deixis in the experimental video-mediated settings, and in the coordination of the teacher / student perspectives for video-mediated lecturing. As an outcome, a deixis-based approach and a framework for spatial design and usability evaluation is presented.

Table of Contents

List of Figures

List of Tables

1 Introduction: Motivation and Focus of the Thesis

This thesis belongs to multidisciplinary studies of Information and Communication Technologies (ICT). It focuses on *interaction space* investigating the formation of shared understanding in collaboration[1].

The motivation of the thesis arises from an increasing adoption of Information and Communication Technologies (ICT) and their impact on the entire workscape; the practical aim of the thesis is to inform the design of the settings that provide resources for effective technology-enhanced collaboration.

Deictic practices refer here to a collaborative situation where the participants communicate their individual points of view to the co-participants in multimodal ways.

Effective communication is understood here as an outcome of the participants' effortless communicative acts, deictic practices, that lead to a shared understanding of their objects of reference which is a precondition for successfully carrying out any joint task. If the participants fail to establish shared understanding, their collaboration is ineffective. To take an example from an ordinary videoconference, it is useless to point with a hand gesture to an object located at the remote site because pointing gesture loses its communicative function to guide the attention of a remote addressee: people there cannot locate the indicated point without related verbal explanation.

The impact that ICT-mediation has on communication and collaboration will be called here *modal fragmentation.*

Taking a multi-methodological approach, the thesis addresses deictic practices in interpersonal communication with the aim to explicate *the human-human interface* as a conceptual structure for workplace design (Table 1). The theoretical background concerning the *role of space* in human communication is in the studies of Architecture, Human Communication, and Human-Computer Interaction. A framework is developed to investigate deixis in a collaborative situation. The conceptual focus of the framework is on deixis whereas the technical focus is on the human-environment interface; thereby, it is possible to investigate the establishment and maintenance of shared understanding in the course of collaboration. Particular attention is paid to a pointing gesture: this is because it makes possible to explore the *intentional* aspects of deixis through the *spatial*

[1] The author's prior papers address particular aspects of this thesis. Joint papers are in-text referenced in acknowledgements to the co-authors' contributions, whereas individual papers can be found in Appendix 1.

coordinates of a communicative situation (e.g. Diessel, 2006; March & al, spatial deixis: Danesi & al. 2000). Furthermore, in concert with verbal communication, it makes possible to explore *cross-modal aspects of communication* in the technology-enhanced settings.

Natural observation, video-ethnography, scenario-based design and usability evaluation are employed. Two case studies are conducted. The first case study, due to refine the framework, is a user test on an exploratory test bed[2] for video-mediated pointing. The critical aspect in it is *effectiveness* of the test bed setting in enabling the communicative function of an across-site pointing gesture. Due to complexity of gestural deixis, evaluation is based on video analysis of the test sessions (Table 6). Another relevant issue for workplace design is *user experience* in terms of (un)helpfulness and (in)convenience of the setting in across-site communication/collaboration. Narrative data is drawn on in the identification of strengths and problems encountered by test participants in the setting.

The second case study, a practical design case, is conducted to evaluate the framework's capacity to inform hybrid interaction design[3]. Evaluation is there based on user ratings of communicative aspects of the setting, narrative data of user experience, participant observation and video data (Table 8). The methodological outcome of the thesis is a deixis-based design paradigm that situates reciprocity of perspectives in a communicative situation.

This chapter is organised as follows: it starts by a brief outline of the technological context change; it then points out modal fragmentation as a design problem; further on, it discusses hybrid interaction design, and to the end, it describes the structure of the thesis.

1.1 Communication, Collaboration and Built Environments

Communication is a joint effort in which the participants focus on a common topic. It has an inherently non-separable nature: when one of its components changes, the whole communicative situation changes accordingly. Conversation is a typical example of collaboration where the participants have complementary roles and shared responsibility for coordinating their perspectives and individual contributions (e.g. Clark & Shaefer, 1989).

[2] The test beds were implemented by Guido Kuehn, and Swen Walkowski from Wiesbaden University of Applied Sciences, Germany.

[3] Hybrid interaction refers to situations where a person is communicating / collaborating at the same time with people some of whom are collocated while the rest are in the ICT-mediated presence.

Table 1 The outline of the thesis

FIELD OF RESEARCH	**MULTIDISCIPLINARY STUDIES OF INFORMATION AND COMMUNICATION TECHNOLOGY (ICT)**		
MOTIVATION	Adoption of ICT innovations > -Workscape diversifying -- -Novel types of communicative situation **Resources for deictic practices in different work environments**	In the video-mediated communication, an across-site pointing gesture loses communicative function **Problem in establishing shared understanding of an object of gestural deictic reference (=what is pointed /referred to)**	-Properties of interaction space for effective communication -in collocated settings -in ICT-mediated settings **Interaction space from modal & cognitive ergonomic point of view**
TOPIC	**THE ROLE OF SPACE** **IN ESTABLISHING SHARED UNDERSTANDING IN COLLABORATION** What are communicative properties of space? In particular, what is the impact of ICT on resources of spatial information for effective communication?		
PERSPECTIVE	**SPATIAL DESIGN (Workplace design)**		
DESIGN ISSUE	**HOW TO DESIGN** **FUNCTIONAL AFFORDANCES FOR EFFECTIVE DEICTIC PRACTICES?**		
DEVELOPING FRAMEWORK 1. to investigate deictic practices → 2. to explain situated deixis → 3. to inform workplace design	---------- drawing on literature in different approaches: ----------		
	>>>ARCHITECTURE -Spatial settings for collaboration -Spatial design principles	**>>>HUMAN COMMUNICATION STUDIES** -Verbal /nonverbal modes -Spatial cognition -Communication as multimodal navigation	**>>>HUMAN-COMPUTER STUDIES** -ICT as facilitator of communication and collaboration
	COMMUNICATIVE SITUATION as a unit of analysis		
THEORETICAL FOCUS (Key concept)	**DEIXIS** **as a multimodal communicative bridge between participant perspectives** >>>		
METHODS:	---------- in the validation of the framework for workplace design----------		
CASE STUDY 1	**Problem case: INEFFECTIVE DEICTIC GESTURE** How to restore the communicative function of an across-site pointing gesture in the video-mediated communication? **To be examined through experimental design & usability testing:** Experimental settings for video-mediated communication/collaboration		
CASE STUDY 2	**Problem case 2: RECIPROCITY OF PERSPECTIVES IN HYBRID SETTINGS** How to design an interface that supports the lecturer in guiding the attention of a multi-site audience? How to design an interface that provides the lecturer visual feedback from a multi-site audience in an effortless (=cognitively ergonomic) way? How to enhance across-site social presence and sense of inclusion at the remote site? **To be examined through a practical design case:** A lecture theatre to be converted for video-mediated lecturing Feedback from test sessions and initial use Observation of emerging practice		
EXPECTED OUTCOME	**NEW INSIGHTS IN THE ROLE OF SPACE IN COMMUNICATION/COLLABORATION** **IN ORDER TO DEFINE** **DESIGN PRINCIPLES FOR TECHNOLOGY-ENHANCED WORKPLACE ->** **A FRAMEWORK FOR SPATIAL DESIGN**		

Language provides an inter-subjective bridge to build a common ground (Clark, op.cit.) at the verbal level for joint efforts: a medium to express thoughts and to share individual views. It is also a tool for navigation as it extends in a symbolic form the spatial and temporal boundaries beyond co-presence: it makes possible to describe past experiences, draft plans for future actions, and negotiate alternative paths through imaginative projections.

Emerging patterns of human communication and collaboration go hand in hand with social and technical innovations. Each great shift in the economic history has produced novel types of spatial formation (e.g. Morris, 1972; Kostof, 1985), and ways of collaboration. A nomadic way of life was characteristic for hunter-gatherers who were dependent on on-site resources. The dawn of agriculture, also referred to as the Neolithic revolution (Morris, 1972), is linked to permanent settlement. Surplus production fostered exchange of goods, trade, and further division of labour, typical of city cultures. Trade and craftsmanship flourished in the Middle Ages; further on, the Great Discoveries, emerging science, and printing press widened the worldviews, extending the horizon of knowledge.

The role of particularly designed built environments became more common in the Industrial Era, along with mass production and modern transport systems; at that time, also major cities started to sprawl beyond a walking distance. A whole range of novel types of building, technical settings and communication media emerged; also design education became established. Furthermore, a shift took place among the economically active population from the primary sector to the secondary sector; accordingly, a workplace typical of the Industrial Era was a factory floor. In the 20th century, a further shift took place towards the service sector; high-rise office blocks, corporate headquarters, open plan offices, and prefabricated building elements represent developments of that era; so do also mass media, and flight connections that intensified cross-continental links.

In the latter half of the 20th century, the rise of computation started to have an impact on the work practices and on the built environments. Architecture as a concept extended beyond its traditional domain to Computer Science discourse where it is applied to systems, software and hardware architectures; also the concept of 'Intelligent building' emerged. The statistical figures illustrate, through the penetration rates in the mobile phone subscriptions and in the use of the Internet, an unprecedented pace of the context change (International Telecommunication Union, 2010; Internet World Stats, 2011):

electronic communication has become commonplace, and Information and Communications Technologies (ICT) have become an integral part of the work settings and recreational resources. A distributed workplace seems therefore to be the contribution of the early 21st century to the timeline of workplace development.

The type of work that is conducted in the ICT-enhanced settings is commonly referred to as knowledge work. Along with ICTs, it is possible for organisations to 'overcome' distance, and carry out joint tasks in virtual teams from geographically dispersed locations. Virtual communities co-exist with traditional communities, and through the social media, people are able to participate, share interests, and socialise with persons whom they have never met in a collocated presence. It is apparent from the diversity of the present-day work settings that the very concept of *workplace* requires an updated definition in the 21st century (cf. SANE project). Due to the rapid context change in the past few decades, the attention focuses in this thesis on the *human-human interface from the workplace design point of view*.

Communication can be seen as *deictic practices* unfolding in a communicative situation; it is by foundation an intersubjective phenomenon, a co-operative (intentional) activity that takes place in an orderly way between two or more participants (>perspectives) in their interaction space when they express their individual points of view bringing them in the *verbal* and/or *nonverbal mode* to the attention of their communicative partner(s).

From the workplace design point of view, communication can be seen as social-spatial navigation: the participants are material bodies in their spatial positions relative to one another at a particular point of time. Space and time therefore provide the basic grid for their collaborative effort *independent of* the type of settings and of the mode of communication.

1.2 Modal Fragmentation as Problem in Remote Collaboration

Communicative situation refers to the spatial distribution of the components of communication/collaboration at a particular point of time. In the Time/Space Matrix, the context of communication is classified into four sub-types:

- same place – same time,
- same place – different time,
- different place – same time, and
- different place – different time (Dix et al, 2004, 665).

In order to initiate communication with one another, people have to establish and activate their connection through mutual orientation. They face, however, particular problem in the ICT-mediated situations: though ICT enables real time communication and collaboration over distance, it also breaks down *natural* human-human interfaces (e.g. Heath & Luff, 1991; 2002; Isaacs & Tang, 1994; Kraut & al, 1996; 2002; Dourish, 1997; Hindmarsh & al, 1998; Hindmarsh & Heath, 2000; Heath & al, 2002; Luff & al, 2001; 2003; Gergle & al, 2004; 2007).

ICT-mediation discriminates between information that will remain within the boundaries of the natural, local information space (=collocated interaction space), and information that can be remotely represented in a digital form. Modal fragmentation and information loss through ICT-mediation therefore constrain the scope of what is possible in a *modal* consideration to get into being in the geographically distributed co-operation (e.g. Goodwin, 2000; Gutwin & Greenberg, 2004; Fullwood & Doherty-Sneddon, 2006; Wickey & al, 2007). As a result, people cannot rely as smoothly on cross-modality, and flexibly switch between different modes as they do in a collocated situation; there they can cue one another in multiple linguistic (verbal utterances), paralinguistic (e.g. intonation, pitch), and nonverbal ways (gaze direction, facial expressions, touch, movements, body language in general).

The problem of 'connecting people' in the geographically distributed communication and collaboration is obviously more complicated than what a commercial slogan suggests. In an ordinary videoconferencing, people cannot indicate an object at the remote site with a hand gesture in an effective way: though the interaction space provides visual information for the remote participants *to see* that that a person is pointing to something, the pointing gesture is ineffective in *guiding* the addressee's *attention* to the object of reference.

Apart from availability of information (e.g. Dourish & Bellotti, 1992; Kuusisto, 2004), the question is about representation: an across-site pointing gesture does not provide over the video link relevant information for the human cognition because the addressee has no means from his/her perspective to infer the *direction* of the pointing hand to the object indicated. As a result, the communicative function of that gesture fails.

1.3 Towards 'Hybrid' Interaction Design

The human-made tools (artefacts) serve individual and collective perceptions and aspirations of wellbeing. The field of design focuses on the interface of humans with the physical and social environment. It acquires knowledge, apart from the findings of

multidisciplinary research, through investigating human practices in the natural and experimental settings. Design has close connections to communication studies, workplace studies and usability studies. It feeds knowledge from the post-implementation research and user tests back to refine designs and to develop new products.

The present-day collaborative/communicative situations involve multiple tools and systems: fixed ones such as building and corridor; movable components such as desk and monitor; organisational components such as enterprise and staff; personal accessories such as pen and mobile phone; information and communication systems and electronic databases; user accounts and firewalls.

The traditional architectural design paradigm is based on human practices taking place in the 'brick and mortar' settings. The technology-mediated way of working is, however, a *hybrid* type of collaboration that calls for a paradigm shift in workplace design in order to cover a workscape that ranges across the traditional and technology-mediated settings.

Though ICT-mediation is in the 21st century an integral part of the work settings and recreational activities, communication studies have not sufficiently focused on the *cross-modal aspects of communication* in the technology-enhanced settings. In order to encompass the differences of collocated and ICT-mediated situations in workplace design *within a single framework*, deictic practices and particularly, their (cross-)modal aspects have to be investigated in order to explicate the requirements for effective deixis in collaboration.

From a design point of view, the key thing is *interaction space* and its *affordances for collaboration;* from a technical point of view, the key thing is the *human-human interface* as a *functional connection*.

In order for individuals to collaborate, their perspectives have to be bridged in communication. Human cognition is by evolution attuned to the collocated spatial and social settings, and to navigation in the co-present social situations (e.g. Flinn & al, 2005). Spatial behaviour can be observed from outside, whereas subjective experiences are unique to each person. As mental contents are situated through the person's materiality, they can be located within a set of spatial coordinates through the spatial position of the person: cognition can be considered in a communicative situation a centre of temporal coordination (and thereby, the origo of subjective experience in reference to the past, presence and future)[4].

[4] Here-Now-I –origo (deictic origo) by Buehler (1934) is the centre of subjective orientation.

An object, for instance a material item, is what it is and where it is, but the *meaning*, attributed (assigned) to it by particular observer is another matter: meaning implies an *intentional* aspect of a triadic nature (Peirce, 1868)[5]: something somewhere appears to someone *as something*, at a particular time.

Deixis is the act of reference, context-boundedness. In navigation, deixis is a functional tool and behavioural strategy that connects individual participants' perspectives / points of view to the context and to the topic of communication (e.g. Clark & al, 1983; Tomasello, 2009; Diessel, 2006).

Shared information space (Kuehn & al, 2007) is the participants' perceptual common ground and interaction space where they coordinate their perspectives for communication. People also regulate the boundaries of their interaction space in multiple ways to make it compatible with their individual and joint purposes. The control of interaction space[6] is closely related to the basic inter/intrapersonal and inter-/intra-organisational questions.

Mutuality is a precondition of communication to flow between individual perspectives (e.g. Nardi, 2005). It refers to awareness (as availability of information) of one another in a communicative situation; in a narrow meaning, it refers to a focused attention to one another.

Reciprocity of perspectives is a precondition of a two-way communication: it enables a role switch between the complementary perspectives (such as the speaker and the addressee in a conversation).

Joint attention is foundational for human communication (e.g. Diessel, 2006): it interconnects the participants' perspectives in their shared information space on a common object.

In a technical consideration, human communication is *multimodal navigation* where the verbal and nonverbal channels are used in a complementary way. To take an example, a pointing gesture (e.g. Enfield & al, 2007) often co-occurs with a verbal utterance complementing it in a cross-modal way; in various situations, it replaces verbal utterances. Children use pointing gesture before they can express themselves in words, and people apply it to inform one another in case they do not speak the same language.

[5] According to Peirce (1868), there is a triadic relation of three semiotic elements: (i) sign, (ii) object and (iii) interpretant. (i) Icon, (ii) index and (iii) symbol again are three different ways of referring to object.

[6] Spatial regulation implies even wider/deeper-going questions of interdependence and sustainability, existence, boundedness, and difference: an adequate response to the context change is the condition for humanity to navigate in its ecology and survive. Learning from prior experiences provides an internal compass for navigation.

Pointing gesture relies on co-presence of the speaker and the addressee. Like sign language, it is conveyed by *visual* information. It is a highly economic/efficient mode of communication because it defines the location of an object of reference in a set of spatial coordinates at the same as it guides the addressee's attention to the referent. However, an across-site pointing gesture loses its communicative function in a video-mediated navigation; as a result, the remote site participants require also verbal information in order to establish shared understanding of the object of reference.

In this thesis, *pointing gesture* exemplifies a modal problem in the technology-enhanced collaboration, and thereby, a workplace design issue. Due to its pragmatic and semantic capacity to steer the attention of the co-participants of communication to the object of reference, it has a key role in an investigation of the formation of shared understanding in the distributed collaboration. Because the practical aim of the thesis is to inform the design of the settings for (distributed) collaboration, the research question is:

- *How to restore the communicative function of a pointing gesture in the video-mediated collaboration?*

The research presented in this thesis is based on the following assumption:

- *Knowledge of the modal[7] conditions of deixis helps to design effective and cognitively ergonomic settings for communication and collaboration.*

A multidisciplinary framework is constructed to investigate a collaborative situation in order to find out

- what the mechanism of (gestural) deixis is;
- how to investigate (gestural) deixis in order to inform design;
- what is required to facilitate (gestural) deixis in the technology-enhanced work settings, particularly videoconferencing, and
- how to design *functional affordances* for effective deixis in the hybrid settings.

Two case studies are conducted to explicate the *context of a deictic reference* in a way that is relevant to workplace design. The first case is an exploratory design due to restore the communicative function of a pointing gesture in the across-site pointing. The second one is a traditional lecture theatre due to be converted for video-mediated lecturing to an audience that involves collocated and remote participants. The outcome of the thesis is an HHI-framework (alternatively, 'locus-modus-topic-focus framework') for workplace

[7] Covering here the terms sensory, functional, operational.

design: it guides the coordination of participant perspectives in order to provide, through the settings, a cognitively ergonomic interaction space for the participants' collaboration.

1.4 Structure of the Thesis

The thesis is organised as follows. Chapter 2 provides a multidisciplinary platform for the thesis: it brings together approaches in the studies of Architecture, Human Communication and Human-Computer Interaction (Computer-Mediated Communication) that help to understand the role of space in human communication and collaboration in a way that is relevant to workplace design.

Chapter 3 describes a selection of relational concepts from the multidisciplinary research literature that help conceptual articulation of a human-human interface in a communicative situation.

Chapter 4 presents a framework for investigating deictic practices in the collocated / technology-mediated settings. It describes the mechanism of deixis providing also deixis-based design principles to facilitate shared understanding in technology-enhanced collaboration.

Chapter 5 specifies a research question and provides an account of the methods and techniques applied in the investigation. It shows how a video-ethnographic method can be integrated with an exploratory test set-up in a compatible way to explore ergonomic aspects of human cognition, and thereby, to guide workplace design to compose functional affordances for shared understanding. Furthermore, it reports the test design, test procedure, and analysis of the data (Case study 1), and a deixis-based method to design an interface for 'hybrid' interaction (Case study 2).

Chapter 6 presents the findings from the two case studies. The findings from the exploratory design (Case study 1) focus on the concept, that is, whether the exploratory design is successful in restoring the communicative function of a pointing gesture in the video-mediated communication. The findings from Case study 2 focus on the design principles, that is, whether the design instructions derived from the framework lead to an effective interaction space for technology-enhanced collaboration.

Chapter 7 concludes the thesis by discussing its contribution, limitations, and relevance. It argues for the proposed framework in workplace design, *the locus-modus-topic-focus framework*, in the facilitation of shared understanding ('common ground') in technology-enhanced collaboration.

2 Field of Research

The chapter brings together Architecture, Human Communication Studies, and Human-Computer Studies to describe the *role of space in human communication and collaboration.* The topic is presented from five different angles, starting from the spatial order of the settings, moving to spatial cognition in orientation, use of space in the social encounter, negotiation of a meaning, and finally, communication over a distance. The aim of the chapter is to provide a multidisciplinary view on the question how *shared understanding* is established in face-to-face communication.

2.1 Research in Human Communication and Collaboration

Human communication is an intersubjective (>multi-perspective) phenomenon, intentional action made manifest in the communicating participants' spatial behaviour. It is deeply rooted in the evolution of human species and cultural development (e.g. Flinn & al, 2005; Elias, 1991); it involves multiple sensory modalities and channels of communication, including speech as an audial expression conveyed by sound, body language as a visual expression conveyed by light, and touch as a haptic expression in a physical contact. In a face-to-face situation, communication unfolds as *verbal and nonverbal deictic practices*.

In this thesis, an investigation of the deictic practices is grounded in the multidisciplinary theory of the role of space in communication through the following fields:

1. Architecture focuses on human communication and collaboration from the *spatial design* point of view with the aim to facilitate communication/ collaboration *through the settings*. In a design, functional and social-cognitive requirements of a joint activity are translated into a spatial layout. Chapter 2.2 describes therefore a selection of spatial articulations that have been developed in the field of Architecture.
2. Cognitive Psychology and Cultural Anthropology provide a dual approach to communication: Cognitive studies focus on *the instrumental role of the human mind* in *the social-spatial navigation*. Social Sciences emphasize the collective aspect of communication as real-life social practices with their *cultural* and *topical regularities/ differences*. Therefore, Chapter 2.3-2.5 takes a cognitive approach to *subjective experience* whereas *intersubjective aspects* of communication are illuminated mainly through an ethnographic lens.

3. The context of ICT-mediated communication differs in a *modal* consideration from the context of natural (face-to-face) communication. It is therefore necessary to take into account *the role of communication tools* (>Information and Communication Technology). Human-Computer Interface (HCI) addresses the human-computer dyad in *information processing*, and Computer-Mediated Communication puts the emphasis on *the interface of communicating people through a technology-mediated connection.* Therefore, the articulations in Chapter 2.6 focus on the *functional affordances* of the settings for interpersonal communication and collaboration.

2.2 Organising Space for Human Activities

Space is, as Kornberger & Clegg (2003) put it, the precondition for processes of organizing: the *primary frame* in which physical and mental activities take place. From the composition point of view, space accommodates the participants of communication as their foundational base (>pervasion, position, relation, and function). Erecting buildings is externalizing human intentions: composing plans and implementing them in a material form to provide settings for human purposes (e. g. Morris, 1972; Benevolo, 1977; Hillier & Hanson, 1984/1989; Kostof, 1985; Bloomer & Moore, 1977; Oksala, 1994). The built environments reflect human action, interaction and social order in a symbolic way, to put it in Heideggerian terms,[8] 'dwelling' in the world.

Architecture operates through metric and relative terms such as volume, proportion, relation, connection, combination, modification and multiplication.[9] It combines human experience/action, communication/interaction, and material components in a *spatial layout* of the settings. The socio-cognitive aspects of space appear in its use (e.g. Hall, 1960; Goffman, 1964; Kendon, 1990; Hillier & Hanson, 1984/1989; Nova, 2003).

The history of architecture is also a history of theorizing in the human-environment relationship. The following articulations of space emphasize different aspects that have to be taken into account in the design of the settings.

In the Vitruvian architecture, human body with its proportions[10] provides the basic yardstick for practical design (Figure 1). An accentuated role is given there to symmetry (>mirror image), and harmonious proportions (Figure 2): apart from the structural

[8] E.g. Heidegger's discussion in 'Building, Dwelling, Thinking' (1971).
[9] Ching (1979) provides a pictorial grammar of different ways of organising space, such as axis, symmetry, hierarchy, repetition, etc.
[10] Body-based measures are still commonly used, to foot as an example.

firmness *(firmitas)* of the building, the spaces have to be well proportioned, and mutually organised with regard to the practical purpose of the building (*utilitas*). Aesthetic experience is in the Vitruvian view understood not only in terms of visual appearance but also as harmony and wellbeing with/in the environment (*venustas*). Ching's (1979) classification of the physical, perceptual and conceptual order[11] is a more contemporary articulation that bears a resemblance to the Vitruvian criteria.

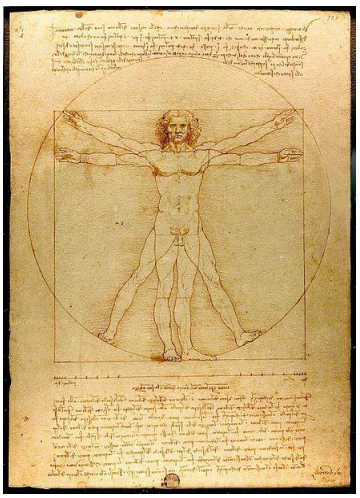


Figure 1 The Vitruvian Man depicted by Leonardo da Vinci[12]
Human body provides an ergonomic yardstick and reference point in practical design.

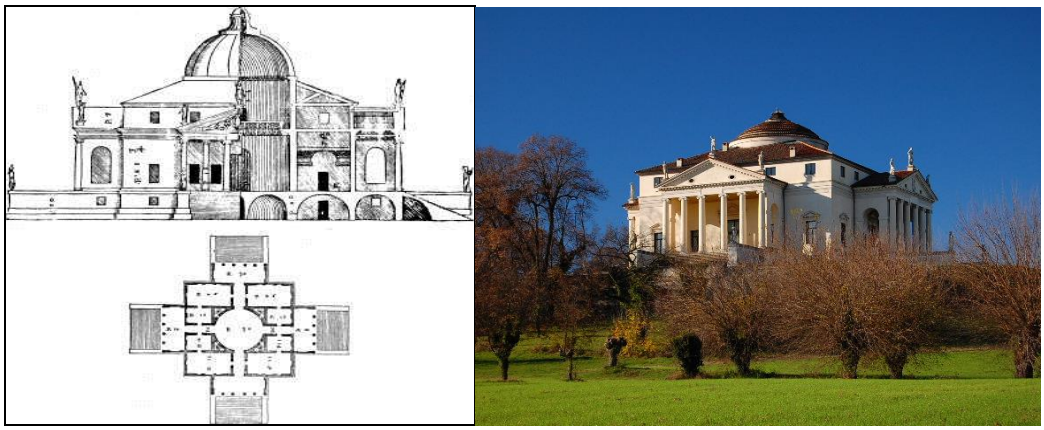


Figure 2 The design of a Renaissance villa (La Rotonda) by Palladio
The composition is based on symmetry and balanced proportions.[13]

[11] For other triadic conceptualizations of space (in geography, sociology, and environmental psychology), see e.g. writings by Tuan (1974; 1977), Lefebvre (1991), Harvey (2004). In HCI and CSCW, e.g. Fitzpartrick & al. (1996); Harrison & Dourish (1996); Ciolfi (2003), and Dourish (2006).

[12] Available at http://it.wikipedia.org/wiki/File:Da_Vinci_Vitruve_Luc_Viatour.jpg (Accessed 05.07.2011)

[13] Available at http://en.wikipedia.org/wiki/Villa_Rotonda (Accessed 08.03.2011)

Also Rasmussen (1959, 9) regards architecture as a functional art considering utility as a decisive criterion in judging it.

Lynch's (1960) writings on the image of the city focus on the *orientation* in the built environment: Lynch asked people to describe the way from a place to another place using projective techniques to study mental images of the city. According to his findings, five elements in the cityscape seem to function as tools of articulation and navigation. Among them, *landmarks* are salient features of the environment that people easily recognize and use as points of reference. *Nodes* are focal points of a gathering nature, such as a railway station is a junction of different modes of transport, or, a market place is the heart of a medieval city. *Paths* are lines wherefrom people perceive the environment during their navigation. *Edges* are demarcating lines between *districts* that have some common feature characterizing them.

A different articulation of space can be seen in Alexander's (1977) *'pattern language':* Alexander developed with his team a number of interrelated patterns that range across different scales of environment. Each pattern provides a 'problem definition' within a wider (practical and spatial) context, and a 'solution' in terms of *how* to achieve it. Compared to strict design rules, patterns[14] give in design more flexibility for situational variation without losing their gist.

Hillier & Hanson (1984) pays attention to what they call the *social logic* of space. As an example, a house is an enclosure that demarcates not only an interior from the surroundings but it also arranges social relations between the inhabitants ('insiders'), and the rest of people ('outsiders'). Similarly, the city walls in the history used to protect cities from an external threat, and to *enable* and *control access* through the gates. *The settings* thereby constrain and regulate social interaction. In such consideration, space is a tool of *control*.

If we bring together the above views, they all describe the *interface* of the human being with the environment through *the settings*. In such consideration, Vitruvius defines general design principles, and provides an ergonomic yardstick for practical design. Alexander's patterns and Lynch's nodes have a gathering nature. It makes them helpful tools in the articulation of a human-human interface in a communicative/collaborative situation. Alexander turns the attention to the topical and contextual scales. Lynch includes in the spatial articulation a dynamic aspect by focusing on an individual perspective on the environment, and thereby, subjective experience in/of navigation.

[14] E.g. Borchers (2001) applies a pattern approach to Interaction Design.

Hillier and Hanson address the built environments as an arena of social interaction (> social-spatial interface).

As Architecture is closely intertwined with the *use of tools*, the built environments reflect the technological history. Both the history of the cities and the recent statistics show how cultural development and technical innovations have diversified interaction spaces for collaboration. Computer-mediated connections provide a controlled access between people over distance, and social media sites extend interaction spaces to virtuality. Virtual objects are not embodied but exist in the electronic form. In order to have their existence and operational capacity, they require embodied infrastructure (='hardware' in the form of input, output and energy supply mechanisms). It is therefore no wondering that architecture as a term is used also beyond its traditional realm.

Reflecting the increasing complexity of our environment, the contemporary architectural discourse ranges from the ecological to conceptual and deconstructivist views. To take two examples, Pallasmaa accentuates *multisensory* architecture opposing the prevailing dominance of vision (Pallasmaa, 2005, 16); he also speaks for Architecture as an art of being in the world[15] (>human-environment interface):

> *"…our environment is part of our identity to the degree that we have no identity detached from situational factors. There is no identity without external attributes just as there is no human consciousness as such without a specific content"* (Pallasmaa, 1991, 91).

In comparison to it, deconstructivist architecture (Figure 3) appears highly exploratory-expressive, at times purely conceptual: it is testing, even crossing conventional boundaries. It is based on a much more *complex order* than is the case with the classical ideas of symmetry, balance and harmony. Its interpretation of the human-environment interface therefore includes not only a gradual transformation but also an abrupt change, tension, conflict, unexpected option.[16] Through colliding forms, intersecting lines, fragmented shapes, and transparent surfaces, it seems to reflect a context where co-presence and virtual presence intersect.

[15] Heidegger discussed being-in-the-world in his writings (e.g.'Building, Dwelling, Thinking', 1971); Norberg-Schulz (1980) published a book entitled Genius Loci: Towards a Phenomenology of Architecture.

[16] Deconstructivist architecture can be seen for instance in Frank Gehry's Zaha Hadid's, and Daniel Liebeskind's works.

Figure 3 Deconstructivist architecture
A complex order can be seen for instance in Walt Disney Concert Hall[17], LA, USA, (left) (designed by Frank Gehry), and interior of Westside Shopping and Leisure Centre[18], Bern, Switzerland, (right) (designed by Daniel Liebeskind)

One possible interpretation of the two above (two more contemporary) views is to regard them as reflections of the interface of the human with the environment. In such consideration, the former emphasizes temporal depth and continuity, and thereby, the least changing features in the human-environment interface over generations. The latter emphasizes the complex, fragmentary, multifaceted, and often heavily conflict-laden contents of our being-in-the-present-day-world: uncertainty and uniqueness, instantaneity and change. Thereby, it reflects the diversity and complexity of the human-environment interface in the 21st century.

2.3 Spatial Cognition in Orientation

Navigation is a goal-oriented activity where prior experiences and knowledge of the situational environment are drawn upon. Perceptual and Cognitive Psychology (Tversky & al. 1999; Tversky, 2003a; 2003b; 2005; 2008; Tomasello, 2003; Tversky & Martin Hard, 2009; Zacks & Tversky, 2001) provides a view on human cognition as a tool of spatial and social navigation.

When people orient to particular destination, they have in their mind a 'bigger picture' of the area in case they are familiar with it; otherwise, they employ information from external sources such as a map and a navigator or instructions by other people. Tversky (op.cit.) describes spatial navigation as a parallel of spatial cognition and action in space.

[17]Available at http://en.wikipedia.org/wiki/File:Image-Disney_Concert_Hall_by_Carol_Highsmith_edit.jpg (Accessed 08.07.2011)
[18]Available at http://en.wikipedia.org/wiki/File:WestsideInterior.jpg (Accessed 08.07.2011)

Mental images provide, however, no accurate representations of the area[19]. They segment the way to the destination into sub-goals and lines that connect them. The sub-goals then function like salient landmarks[20] along the way. In the course of navigation, the value of a mental estimation is continuously adjusted by coupling the mental images of the area with the situational information of it.

Three different perspectival systems are involved in navigation (e.g. Tversky et al. 1999):

- the person as a (material) spatial entity
- egocentric orientation in the situational context[21], and
- allocentric positioning of oneself in the environment.

The coordination of the participants' contributions is necessary for all joint efforts (e.g. Rosenberg & Sillince, 2000). People navigate not only in the physical but also in the *social* space[22]: they provide information to one another; they form *joint attention*[23] with others, they form shared intentions, and they learn from one another (e.g. Tomasello, 2009).

Shared intentionality is considered foundational for a social effort (Tomasello & Carpenter, 2007), and *reciprocity of perspectives* for socially organized interaction (Schutz & Luckmann, 1973). In the functional consideration, space is the foundational grid where the individual perspectives are coordinated to provide mutually *compatible perspectives* to communication and collaboration.

Apart from the compatible spatial perspectives, communication (as social navigation) requires *relative abilities* and *socio-cognitive competences* (Flinn & al, 2005), among them, empathy and theory of mind (understanding that the other has his/her own subjective experiences), working memory, attentional control and executive functions.

Furthermore, communication requires quick and efficient processing of social information; multiple order reasoning; imagination, fantasy, and creativity; in addition,

[19] Systematic errors have been found in the research in spatial cognition to be a natural consequence of normal perceptual and cognitive processes; they include, among others, distortions of distance and direction (For more, see Tversky's studies in spatial cognition.)

[20] Note similarities between Lynch's and Tversky's views regarding navigation strategies.

[21] Compare with Buehler's (1923) Here-Now-I coordination system of subjective orientation.

[22] There is the inner, psychological reality of a person, the actual world around her, and what Winnicott (1971) calls 'potential space' as the locus of 'transitional phenomena', and thereby, of cultural experience.

[23] According to empirical findings in developmental psychology, intersubjectivity develops in early ontogenetic life in the caretaker-child interactions: shared attention and intentions co-occuring such interactions become internalized and embodied (Flender & al, 2009, 211). Gaze following is transformed in the child's early cognitive development by *shared intentionality* into *joint attention*, social manipulation into cooperative communication, group activity into collaboration, and social learning into instructed learning (Tomasello & Carpenter, 2007).

communication requires an ability mentally to construct and manipulate a range of potential social scenarios, to anticipate and influence social interactions with others, and to anticipate the social strategies of the others. Also an ability to detect deception, and mentally to stimulate and evaluate potential counter-strategies is required (ibid.).

The above views on cognition appear to be congruent with Buehler's (1934) coordinate system of subjective orientation, and with the ecological perceptual view (Gibson, 1977; Gibson, 1979) on inseparability of human / environment. Furthermore, they are congruent with the situated view on cognition (e.g. Roth, 2004) that regards a real-life situation as the informational basis on which the cognitive contents emerge in the flow of events (within a particular frame of interpretation) as a unique composition. Also the embodied cognition view (e.g. Merleau-Ponty, 1945/1995) holds that there is no cognition without material foundation for its emergence.

In the light of the reviewed literature, the spatial position of a person is of key importance in the analysis of communication from a workplace design point of view: it is a coordination site that provides the spatial coordinates in relation to the surrounding human and non-human objects (>individual spatial perspective); at the same, it provides situational information for the person's (sensory caption and) perception. Furthermore, it maps the person's situational cognition and intentional orientation (>cognitive perspective parallel to the spatial one) in the course of his/her social-spatial navigation.

2.4 Use of Space in Social Encounters

The following account of human communication draws on the findings in the social-cognitive studies. It illuminates how people *use space to regulate the contact with other people*. Spatial control occurs in multiple forms, and at different levels of a sense making process.

Social cognition refers to the mental representations and processes that underlie social judgements and behaviour (Smith & Semin, 2007; Gabora, 2008).

Elias (1982: 1991) describes social-cognitive patterns that guide conduct and polite manners as culturally internalised self-control. They have developed over time to increase predictability in the social encounters: they provide basic common structures so as to help avoiding unnecessary conflict and misunderstanding in communication – that is important

particularly in the situations where people are not familiar with one another[24].

Goffman focuses on the *situation* (1964) (For more, see Chapter 3.1), on the *roles* (1956/1959), and the *frames* (1974) of social interaction. He emphasizes communication as a performance that unfolds in the physical and social settings: people are, as if, on the stage of a social theatre where they have to look after impressions they make on others. In the social interaction, the *face* counts a lot: in the *front stage*, people have to focus more on the role, whereas in the *back stage* they can relax from keeping up appearances.

Kendon (1967; 1990) addresses gestural communication; his gesture studies illuminate social encounter as a spatial performance. The 'choreography' of particular spatial behaviour has a communicative function embedded over time as a *pattern* in a cultural convention: for instance, when meeting, people shake hands, and when leaving one another, they do the same again. A face-to-face type of situation can, in such consideration, be seen as a *multimodal spatial performance* where the *relative position, distance, orientation (direction)*, and gestures of the participants (and non-participants) play a communicative role.

Mutual distance is a control of interpersonal communication (e.g. Hall, 1966/1990; Nova, 2003; Kraut & al, 1996; Kraut & al, 2002); if it is too long or too short, it has negative impact on communication. Distance is also used to communicate difference in the role, and in the social position. Hall describes different types of zones around a person, starting from intimate space to personal, social, and public space[25].

Access control is a spatial way to regulate a social distance. (e.g. Morris, 1972; Kostof, 1985). Building a wall means at the same barring access, that is, creating an obstacle that stops people walking through it, whereas building a bridge provides a connection that did not pre-exist. A door not only allows access but when locked, it disables it. A window provides a visual connection while at the same it separates a space by transparent material. In an office, the traffic lights indicate availability of a space and a contact with the person who occupies the space. Scheduling is a temporal way to control access employed particularly in the office environments.

New forums for communication and collaboration bring along new types of condition for collaboration and new controls of distance. The control mechanisms of interaction space involve, among other things, spatial controls such as walls, doors, shape, size; temporal controls such as pace, schedules, time tables, deadlines; social(-cultural) controls such as

[24] For more regarding the topic, see for instance Freud (1930), Discomfort of Civilization.
[25] There are cultural differences among others in an appropriate interpersonal distance.

laws, norms, convention, etiquette; behavioural controls such as access/exit, attention/avoidance, gaze-contact/aversion; talk/pause; technical controls such as automata, systems, standards, protocols, and digital controls such as user identification, and passwords.

Similar to distance, *relative orientation* (e.g. Kendon, 1967; Nova, 2003) has a regulatory role in the interpersonal communication. A gaze contact is typically taken when thanking or greeting someone, whereas staring at people is avoided. It shows that gaze contact plays an important role in the context-appropriate behaviour[26]. Turning the back to a person tends to signal avoidance.[27] Gaze avoidance has also been identified as a means of 'taking distance' for a cognitive task, as if gaining own space for reflection (e.g. Markson & Paterson, 2009).

A number of classifications of space have been presented along the *public-private* continuum. Depending on the classification, there may be one or more intermediary positions: Harrison & al. (2003) for instance classify private, privileged and public spaces. The privileged space implies an access control condition, such as a membership card for a club, and a password for an intranet. Space is regulated also on a gender basis, to take Ladies/Gents as an example.

Inclusion/exclusion and involvement (engagement) imply both spatial and social aspects of communication. The degree of inclusion varies in a communicative situation between a participant, onlooker, over-hearer, and passer-by. Among the 'insiders', the degree of participation varies between the active and the passive participant. A tracker is involved in an asynchronous way in a communication (Rosenberg & al, 2005). Ownership of a space represents socially legitimized control of a space and related behaviour within it.

As a summary of the above, human communication is governed by *visible* and *invisible* constraints: on one hand, spaces are implemented for particular purpose(s); the layout of the building[28] – to take a an amphitheatre as a poignant example - guides and regulates the conduct in it; on the other hand, people interpret a building as a setting of particular activity/ies and *context-appropriate behaviour*. Their spatial conduct then regulates the interpersonal contact according to a person's situational interpretation / motivation.

[26] Apart from physical distance and direction, gaze communicates also emotional distance as it may be welcoming or refusing a contact.

[27] Interestingly, many linguistic images reflect control aspects of space, for instance in English 'think highly of himself', look 'down on someone'.

[28] Foucault discussed (e.g. in Discipline and Punish, 1977) mechanisms of control.

2.5 Use of Language in Negotiating Shared Meaning

Different theoretical views on language articulate the interpersonal interface in a way that is informative for workplace design. They are presented in the following paragraphs.

In a cognitive-functional consideration, human beings conceptualise the world in terms of certain configurations such as space and time, force dynamics and causality, perspective and attentional distribution (Talmy, 2003, 5, 15-46). A perspective and attentional distribution imply the position of the person, and the position of the object observed, and thereby, *orientation (>direction)* in the context of communication.

Tversky (2008, 441) regards language use as *a multi-level phenomenon*: language is structured on multiple levels of organisation such as sound, meaning and discourse. These levels *cooperate* and *interact* but are *not* completely *reducible*. Research findings also suggest that the same perceptual-motor mechanisms that underlie *production* of speech also underlie *perception* of speech (ibid. 440).

Where Tversky (2008) points out the *physical, cognitive* and *social* dimension of language use, Buehler (1934) brings together the *spatial, temporal* and *perceptual (>modal)* aspects through the Here-Now-I -coordinate system of subjective orientation: when a person refers to an object in his surroundings by a word, or, in a nonverbal way for instance by a hand gesture, s/he is (where s/he is at that very moment) the origo of a *deictic field.*

Symbols are for Elias (1991) the 'fifth dimension': Elias points to the biological evolution and the social development in the background of symbols, language, and accumulation of human knowledge. Symbols are in his view tangible sound-patterns of human communication, made possible only by the precondition of the human vocal apparatus.

In developmental psychology and in the usage-based approach to language acquisition, symbolic language is considered to be a derivative of *shared intentionality* (e.g. Tomasello, 2003; Diessel, 2003; 2006; Flender & al, 2009, 211).

Kravchenko's (2007) bio-cognitive view emphasizes the *connotational* nature and *orientational* influence of language. Language is in such view *adaptive behaviour* of an organism that involves a system constituted by signs of signs. Representations are sign entities whose biological function is that an organism, by interacting with them, adapts to the medium by managing information; they are mental structures born of experience, and they function as a storage facility for knowledge. The essential feature of communication

is that an organism has the intention to modify another organism's behaviour in a particular way.

Research in cognitive linguistics recognizes *interrogatives* and *demonstratives* as part of the basic vocabulary of different languages (Diessel, 2003; Diessel 2006). Interrogatives are tools for symbolic parcelling of the world into conceptual components. Demonstratives have *an orientational role* as they are primarily used *to focus the addressee's attention to an object* (Diessel, 2006; Clark & al, 1983; Lloyd McBurney, 2002). For instance, if a person points to an object uttering: 'I mean *that* book, not the other one', only a person who is present in the speech situation can identify the object referred to. Demonstratives thereby combine pragmatic and semantic aspects of discourse.

Nonverbal communication is an integral part of a face-to-face speech situation. Goffman (1964, 133-136) considers speaking as part of a complex human act. Kinetic features of speech behaviour include *facial expression, gaze orientation, hand gesture,* and *change of posture* and *movement*. Psychological features of speech behaviour are expressed in *affect displays* such as facial expressions, sounds, gestures, and movements (=body language) (e.g. Darwin, 1872).

From the *practical* point of view, language serves humans as a means of *symbol function*: it extends the here-and-now experience by providing people symbolic tools for 'handling' what is elsewhere. It helps *to bridge individual perspectives* and share experiences; it provides a tool for reasoning and drafting plans together with other people.

Common Ground approach (Clark, 1996) in cognitive linguistics focuses on the question how people create *shared understanding* in a collaborative way *at the level of language* (For more, see Chapter 3.4). Roth's (2001, 2004) and Gill's & Borcher's (date not available; 2005) work on gestures and body language extend the view on mutual understanding towards an extra-linguistic domain of communication.

From the organisational point of view, the flow of communication unfolds in a social encounter as interplay of the participants' contributions (For more, see Chapter 3.5). Communication is therefore *a joint effort* made possible only if it is jointly contributed and coordinated by its participants. Social arrangements guide the contributions because each utterance must be referred to the state of ongoing talk that is sustained through turn taking (Goffman, 1964, 136). Turn taking and floor control serve the coordination of verbal communication.

For the procedural management of a conversation, *cues* must be available for requesting and giving up the floor. Cues inform the speaker about the stability of the focus of the addressees' attention; they provide also feedback whether the audience understands what s/he is saying (ibid.). *The scope of the topic* and the *focus of the topic* (Sillince, 1995) are structural controls to maintain the coherence of discourse (e.g. a conversation).

The question how speech situation is interpreted from the participant's point of view at the time of talk has to be taken into account in a social-cognitive analysis of communication (e.g. Akman, 2000). Context model is a mental model in episodic memory (v Dijk, 2006); it controls and explains aspects of interaction that cannot be analytically reached through other type of approaches to text and speech communication. Context model[29] therefore provides a theory of *relevance* and *situational appropriateness* (ibid.).

In conclusion of the presented views, language use is a *social-cognitive, modal and spatial phenomenon* that opens up to the syntactic (grammar, 'configuration'), semantic and pragmatic dimensions. Though some of the above described views clash with one another in some respect, they all share space and time as the basic coordination system. Language can be seen as a system of signs *used* by humans *for their joint navigation* in the world. In that consideration, language is a *tool* among other tools, an artefact developed over human generations *to enable a symbolic common ground* for interpersonal communication and collaboration. Language use can be located ('mapped') in a set of spatial coordinates through *the speaker's spatial position in the speech context.* In a procedural consideration, interrogatives and demonstratives have an *orientational function similar to pointing gesture in the nonverbal communication*: they all serve the search of a referent in the speech context.

2.6 Use of Tools for Communicating over Distance

Computer-mediated communication (CMC) turns the attention from the face-to-face communication to the communication and collaboration over a distance. It addresses media configuration and usability issues such as effectiveness, efficiency, and user acceptance. Other foci of interest are access to and control of the resources, the features of interaction spaces and their different affordances for communication and collaboration. Human-Computer Interaction (HCI) focuses the human-machine interface in its structural, functional, cognitive and practical aspects (e.g. Winograd & Flores, 1986;

[29] Compare with Goffman's framing of experience.

Rogers & al, 1994; Dix & al, 2004; Nardi, 2005).

Usability of the tools is of key importance in practical design. Affordance[30] is an interface concept for an actor with his/her environment. Norman (1988) introduced the concept to the Human-Computer Interface community that readily adopted it for design and usability studies. Over the years, different authors have contributed to a rich affordance discourse in design literature (For more, see Chapter 3.7).

Turner (2005; 2008) addresses the role of *artefacts* as part of, and in relation to, human action. He shows how, despite their very different approaches, Heidegger and Ilyenkov equate *context* and *use* arguing that we understand the world in terms of use. From a holistic point of view, the world is a referential whole,

- a totality of interrelated pieces of equipment (such as for instance a hammer);
- a set of purposes, (such as using a hammer for fixing a chair),
- an identity assumed when using it (such as being DIY-carpenter).[31]

Familiarity[32] as 'readiness' to cope with things encompasses the ideas of *involvement* and *understanding*. Involvement refers to the fact of being-in-the-world and understanding to its practical and epistemic aspect (know-how), in the form of tacit knowledge of everyday activities (ibid, in reference to Heidegger).

Activity is (in Ilyenkov's[33] view) the principal expression of how we inhabit the world. Social life is a product of the collective, but experienced by individuals as a set of given rules, practices, tools, and artefacts. Objects acquire their 'significance' through their use. Significance itself is not yet affordance but it makes a thing *knowable;* it has to be attached to the thing through human activity (Turner, 2005, 796).

Virtual/distributed settings have become an integral part of our practical contexts. Their differing affordances for communication and collaboration are discussed widely in design literature (e.g. Nardi, 2005; Gaver, 1991; Gaver, 1992; Gaver, 1996; Isaacs & Tang, 1994; Heath & Luff, 1991; Kraut & al, 1996; Kraut & al, 2002; Hindmarsh & al, 1998). Such settings heavily constrain multimodality of synchronous communication and collaboration. In addition, the establishment of *mutual orientation* becomes an obvious

[30] The notion of affordance was coined in ecological perceptual psychology by Gibson (1977; 1979).
[31] DIY= do it yourself
[32] Familiarity is a notion that Heidegger used in a very deep meaning referring to how people cope with things in the world; familiarity has a central role as our coping mechanism with the environment.
[33] Ilyenkov's classification of non-material phenomena into mental and ideal phenomena appears to have a resemblance to Popper's ontology of three worlds: World 1 (nature), World 2 (subjective experience) and World 3 (culture).

problem due to incongruence of the participants' perspectives (e.g. Hindmarsh et al, 1998; Hindmarsh & Heath, 2000). Mutual / joint orientation is, however, particularly important in communication as it provides the foundation for mutual/shared understanding.

Social presence theory (Short & al, 1976) contrasts face-to face communication with technology-mediated communication in terms of awareness of the other part of communication. It classifies media along a continuum from high degree of social presence (such as face-to-face communication) to low degree of social presence (such as text-based communication). It holds that the communication medium has to be, in order to be effective, appropriate to the type of the joint task and involvement of the participants. The theory thereby implicitly points to the complementary roles and turn taking in communication: awareness of the other part's actions is crucial for initiating own contributions in a smooth and seamless way. Social presence theory is also relevant in terms of how people present themselves, and how they are interpreted in the social media.

The school of *media naturalness* in the CMC community emphasizes natural-likeness of communication (Kock, 2005): it holds that the medium effects on a cognitive effort, physiological level of arousal, and ambiguity of communication. It considers a capacity to convey speech an important yardstick of media naturalness. The theory points also to the flexibility of humans to adapt and make the best use of available resources; thereby, some of the negative effects of media are compensated. The topics of media naturalness overlap those of cognitive ergonomics.

Media richness refers to the modal capacity of the medium (e.g. capacity for immediate feedback). Proponents of media richness view (Daft & Lengel, 1986) hold that the more personal and rich the medium is, the more effective it is for communication. Yet, people have also been creative in developing compensatory techniques, to give emoticons in the text-messages as an example of social cuing.

Social information processing theory (Walther & Parks, 2002) suggests that developing trust might take more time in online communication compared with collocated communication.

The above views highlight the complexity of the interpersonal interface in the ICT-mediated settings. The *affordances* for synchronous communication and collaboration depend not only on the participants' capacities but on the settings as well: the technical channel(s) constrain(s) the *mode* of communication and collaboration. Compared with the

natural face-to-face communication, distributed settings differ from the collocated settings in terms of a functional connection (= natural/artificial), and a frame of reference (=spatial layout of the resources, perspectival in/congruency, visual frame/scope)

ICT-mediated settings complicate the process of mutual orientation and joint attention, and thereby, the establishment of shared understanding in virtual communication. Therefore, particular attention has to be paid to *modal and cognitive ergonomic aspects of communication* in the design of the settings for (distributed) collaboration.

2.7 Summary

The chapter described, in brief, approaches to human communication and collaboration in a multidisciplinary research literature. A human-centred approach to design (e.g. Vitruvian architecture) considers the human-environment relationship as a dyad ('mirror of nature'); accordingly, the phenomenological interpretation considers architecture 'the art of being in the world' (e.g. Pallasmaa). Deconstructivist architecture gives expression to an abrupt context change, and to overlapping contexts.

In an analytic consideration, the articulations of space as the 'primary frame' provide tools for the 'quantitative fit' (in terms of volume, proportions and functions), and for the 'qualitative fit' (in terms of ergonomic, aesthetic, social relations). In practical design, human body provides an ergonomic yardstick. Lynch (1960) highlights the role of *reference points* in a person's navigation of the built environment. The reviewed literature also provides a set of *general* (Vitruvius, Ching, 1979), and interrelated *topical* design principles (Alexander, 1977).

Cognitive studies focuses on the *mental representation/orientation*, and *intentional action* (>*navigation*). *Patterns* of social interaction (Kendon, 1964; Kendon, 1990) characterize *social conventions* in a way that is applicable to spatial design. *Spatial cognition* (Tversky, op.cit.), *perspectival systems*, and *framing systems* (Buehler, 1934; Goffman, 1974; Clark, op.cit; v Dijk, 2006) are helpful conceptual tools in the analysis of a communicative situation, and in the design of the settings for collaboration.

The field of design interrelates the individual perspectives (through the settings) with the aim *to facilitate individual and joint human activities*. The reviewed literature shows that

the 'modal fit'[34] and the 'cognitive fit' are particular challenges in the coordination of interpersonal interface through the technology-mediated settings.

My conclusion of the literature reviewed in the chapter is the following: A communicative situation has to be addressed *through complementary approaches* in order to cover *phenomena that are relevant to workplace design.*

In order for a framework to be robust enough to guide the design of the settings, independent of whether collocated or distributed collaboration is in question, *the interpersonal interface* has to be described as a conceptual structure. Therefore, *the human-environment interface* has first to be conceptualised (Figure 4). Only then we are able to describe the interpersonal interface in a communicative situation in a way that is relevant and applicable to workplace design.

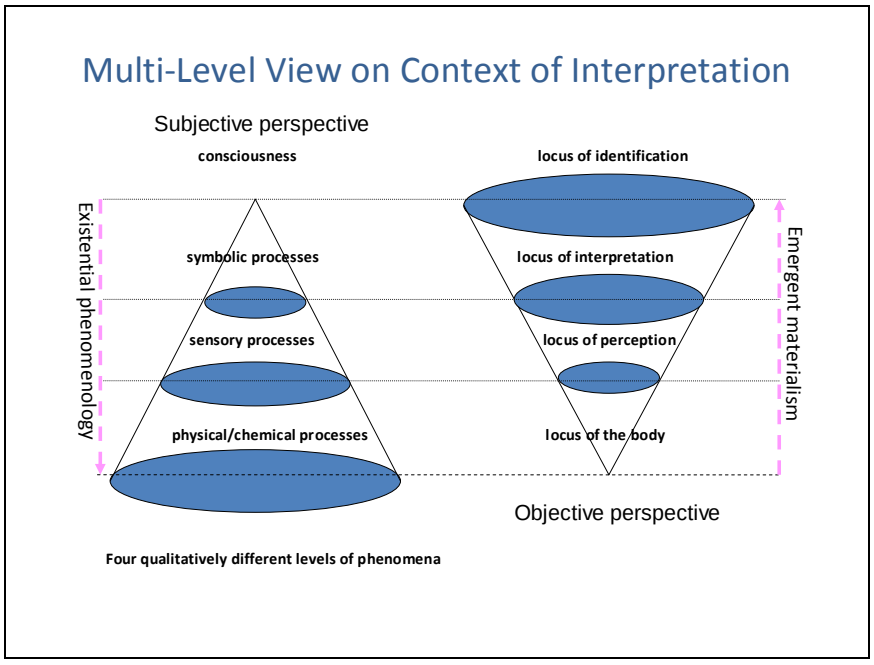


Figure 4 The human-environment interface as a multi-level structure
The figure illustrates complementary views on the human-environment interface.
Existential phenomenology considers subjective experience semantic layer in human existence (Rauhala, 1993). That interpretation brings forward the uniqueness of individual perspective. Emergent materialism provides a complementary view, a diversity of individual perspectives (e.g. Popper's three world's ontology); in such consideration, individual perspectives can be seen as isomorphic structures with different contents. As a frame of interpretation, they have a common ground because each of them is a human perspective that has developed within a cultural context (in interaction with other people).

[34]Referring to the connection (functionalities and sensory-motor capacities of the participant of communication).

The following notions (in the reviewed literature on the role of space in human communication) are all interface concepts. Due to their relational nature, they provide helpful tools in the conceptual articulation of the human-environment interface, and the human-human interface in a communicative situation (For more detail, see Chapter 3):

- communicative situation, experience, attention, deixis, common ground, cross-modality, and affordance.

In a methodological consideration, a cognitive approach addresses human experience and intentional action from the *individual actor's perspective*. It is therefore relevant/applicable to the design of test settings to explore the use of an experimental system from a cognitive ergonomic point of view. Buehler's (1934) coordination system of subjective orientation, Tversky's (1999) spatial cognition and three perspectival systems in navigation, v Dijk's (2006) context model in episodic memory, Goffman's (1975) framing of experience, and Clark's (1995) 'common ground' in a conversation are in a cognitive consideration helpful tools.

Yet, a cognitive approach alone is not sufficient to explain communication in real-life situations; communication implies internalisation and externalisation in individual participants' perspectives, when intentions are 'acted out' in the real-life environment, and experiences internalised from the flow of events (Figure 5). Therefore, communication cannot be fully analysed without taking into account *the situation as a cultural* and *spatial setting* where *social-cognitive practices* take place. Goffman's (1964) view of a social situation as the 'locus' of communication, Kendon's (1990) spatial pattern of a focused encounter, Tomasello's (2003) view on usage-based acquisition of language, and Kravchenko's (2007) view of the connotational and orientational nature of language point all towards observational methods.

Ethnographic observation is a method that focuses on real-life phenomena, and therefore, on communication as a *spatial performance* between two or more *individual perspectives* in a particular *cultural context*. In comparison to a cognitive approach, it emphasizes *intersubjectivity* and *community of practice*. From a design point of view, it maps social interaction as 'choreographies' describing their *regularities* as *patterns*. Thereby, it captures a range of different people's performances as they are encapsulated in *social conventions* in the observed culture (e.g. Geertz, 1973). Ethnographic observation therefore complements a cognitive approach in an ideal way in an investigation of deictic practices. (For more, see Chapter 5.3).

3 Theoretical Background

The chapter presents a selection of notions from a multidisciplinary literature on the role of space in human communication in order to articulate the interpersonal interface in communication. Each notion is concluded from a design point of view.

3.1 Communicative Situation

Though communication can be approached from different angles - taking Burke's pentad of motives act, scene, agent, agency, and purpose as example (Burke, 1945, xv), - it has an inherently non-separable nature. Social situation arises, according to Goffman (1964), when two or more people find themselves in the presence of one another, and last until the next-to-last leaves: it is

> *"…an environment of mutual monitoring possibilities, anywhere within which an individual will find himself accessible to the naked senses of all others who are 'present', and similarly find them accessible to him"* (Goffman, 1964, 135).

Communicative situation was long neglected in the analysis of speech. Yet, Goffman argued in 1964 for social situation in its own right to be an adequate unit of analysis: thereby, he turned the focus towards a more holistic and pragmatic approach to human communication and collaboration, and methodologically, towards ethnography of communication[35]. In his view, it is not possible to analyse speech behaviour such as a conversation without taking into account situational correlates such as the physical settings where it occurs, and the presence of other people at the time of talking. Neither can the study of the behaviour of the speaker be analytically separated from the study of other people present to him/her - independent of whether or not they are engaged in talk. Otherwise properties of their mutual relationship would be left out of analytical attention. In order to describe for instance a pointing gesture, a wider extra-linguistic context where the gesture occurs has to be taken into account; the analysis has therefore to include both the spatial and social settings.

A primary framework involving situations (as events taking place) is, in Goffman's view (1974), the physical world. The second frame is a social framework; it refers to organizing events and connections within the frame of human perspective (ibid.).

[35] Saville-Troike (1982) further elaborates ethnographic current in the 1980's.

Goffman (1964) classifies *gatherings* and *encounters* as sub-types of a social situation. A physical coming together – as it is the case in a collocated gathering - is the ecological huddle for participants *to orient to one another*, and at the same time, away from others that are not involve in the encounter (though being present). Participation in a gathering entails *constraint* and *organisation*: cultural rules organise the behaviour by guiding people to conduct themselves in a context-appropriate way. A more specific type of joint orientation is *encounter*. It refers to two or more people

> *"jointly ratifying one another as authorized co-sustainers of a single, albeit moving, focus of visual and cognitive attention"* (ibid.).

Mutual ratification is needed to initiate communication, as if opening the channel for communication to flow between them. There are particular rules of the initiation and termination of an encounter: talk is socially organised not only in terms of who is speaking to whom in what language, but also as a system of mutually ratified and ritually governed face-to-face action, a *social encounter* (ibid.) Encounters differ in terms of the number of participants and in terms of activity; they also vary in terms of (in)formality.

Goffman's interpretation of a social situation is supported by Malpas (2002, 403) according to whom meaning is realised only in relation to particular settings or 'locales'. Similarly, Hindmarsh & Heath (2000) point out that intelligibility of referential actions is grounded in the activities. Cognition is also *socially situated* (e.g. Smith & Semin, 2007, Gabora, 2008), and an individual perspective implies a *mental context* at the time of communication (v Dijk, 2006).

To summarise the above from a workplace design perspective, communicative situation is the locus where situational information through different communication channels is brought together in individual perspectives, and coordinated at/through different levels/frames of individual / interpersonal sense making process. Communicative situation is, in a spatial design consideration, the spatial layout (distribution) of the settings, participants and objects involved.

3.2 Experience

In Rauhala's existential-phenomenological view (1993), experience is a *semantic layer* in human existence[36]. Human body is a necessary *'instrument' for experience*, whereas the

[36] Also the Three Worlds ontology by Popper distinguishes between nature, subjective experience and culture (Niiniluoto, 1990; Eccles & Popper, 1984). According to Popper, "World 3 is the world of the products of the human mind" (Popper et al., 1977, 449).

person's life situation provides its *factual basis*. Rauhala (1991; Rauhala, 1993; Koskinen & al, 2003) accentuates life situation[37] as an inseparable part of human existence: it refers not only to spatial relations but also to any type of relation between the person and the rest of the world.

Tversky & al. (1999; 2008; Zacks & Tversky, 2001, Tversky, 2005) take a perceptual and cognitive psychological approach to experience. In a perceptual consideration, human body is a multimodal instrument, a '*moving set of sensors*.' The continuous information input is too rich and complex to take in, and therefore, it has to be categorized somehow. The human mind organizes and integrates sensations captured by the sensory apparatus in the stream of information from the surrounding world and of the body's own movements and processes. It demarcates events from the sensory flow articulating them by partitioning mechanism such as wholes and parts, their relations, and functions they serve (Zacks & Tversky, 2001; Tversky & al, 2008).

The object of observation is perceived and interpreted *as* something (=a meaning attached by the person to the object)[38]. Gestalt psychological school of thought points out *figure* vs. *ground* as a basic articulation mechanism for perception (e.g. Koffka, 1935): the human mind demarcates something as a figure from the background, that is, from overall information available for sensory perception. The 'figure/ground theory of articulation' therefore implies that perception is *directed (orientational)* by foundation.

A person's position in a communicative situation is a particular spatial perspective onto the surrounding space. S/he is the origo of that information space. Yet, s/he is not the only origo of perception: the other participants of a communication have their own positions and their own perceptions. Communication is therefore by foundation a multi-centric (multi-perspectival) phenomenon: its participants have to link their points of view at the level of action (=spatial behaviour) and at the level of interpretation (Schober, 1993; Goodwin, 2000) in order to get something in a co-operative way into being.

Perception is not just abstract entities but specific *objects* with their attributes such as shape, size, and components. Correspondence of perceived features and their functions/ behaviours implies information that is required in order for people to learn to organise and plan their behaviours (Tversky & al, 2008; Tversky, 2008; Tversky, 2005).

[37] Along with consciousness and physicality (corporeality)

[38] Peirce (1868) described triadic typology of sign; another classification was between firstness, secondness and thirdness.

Subjective experience and intentional action[39] can be seen as two aspects of the same situation (Figure 5). Spatial behaviour provides situational information for interpretation. In this process, the former is the informative, factual basis for mental representation. Seen from the opposite direction, intentional action is initiated in a top-down fashion and *executed* through motor actions. In other words, human mind is *projecting* intentions in the environment, and realising them through his/her spatial behaviour (and possible use of tools). In that sense, intention is a 'program' or a 'plan', situated and concealed in the mind until being made manifest in spatial behaviour (speech, motor action).

In human communication, distinction has to be made between *human objects* and *non-human objects*. Similarly, a difference has to be seen between *natural events* and *intentional actions*. For instance, if the wind slams the door close, there is a natural cause, whereas if a person closes the door, the question is about the outcome of an intentional action (executed in bodily behaviour) by a human actor.

Von Wright (1998, 142) points out that human action has to be understood *under the aspect of intentionality*. Something being a reason to action (for instance, why a person closed the door) means that the relation between mental and behavioural is not causal but semantic: mental states are *attributed* to the subject based on behavioural phenomena (ibid. 147)[40].

To summarize the above views on subjective experience from a workplace design perspective: subjective experience is by foundation spatial and perspectival: the human body is, in a physical consideration, a sensory unit in a spatial context; therefore, *perspective-taking*[41] is an inherent part of the human-environment relationship (>interface), and an individual perspective is a multimodal interface. The position of the person constrains availability of information from the surrounding world for his/her sensory caption, and the sensory apparatus itself sets further constraints by filtering information that is compatible with it. (The sensory modalities operate through different channels, each of them being capable of processing particular type of information)[42]. Different sensory modalities have different operational radius and scope: hearing covers practically 360° of the person's situational information space whereas the scope of vision is only a limited sector at any time.

[39] Suchman (1987) discusses plans and situated actions in human-machine communication, and Mantovani (1996) social context in HCI.
[40] v Wright (1998) describes in reference to human action three different phenomena (mental / neural / behavioural) and relations (epistemic / causal / semantic).
[41] Note also that personal pronouns define three perspectival positions both in singular and plural form.
[42] Human hearing for instance is a sense that operates within the range of ca 20 – 20 000 Herz.

Furthermore, the human perspective implies a *two-way process* of systemic internalisation and externalisation: subjective experience and intentional action (executed through a person's bodily behaviour in the spatial settings).

3.3 Attention

Diessel (2006) recognizes attention as the most basic function of human communication. Capacity

1. to render a figure from the background,
2. to focus on, and
3. to have personal relevance

are crucial for all human orientation and communication.

The situational information from the surroundings of a person provides the informational basis for his/her situational awareness of it. In that sense, the person is origo of a perceptual field. It is not possible for the human mind to focus on everything at one go but attend only to some part of it: something in the field appears as *relevant* and *salient* enough to catch attention as a *figure*[43] demarcated from its background.

The cognitive approach holds that people have direct access only to sensations, which are integrated with memories to build up symbolic representations of the environment and its potential for goal-oriented action (Gaver, 1991, 79).

Attention implies that human cognition is in *intentional* relation to the world; it is foundationally *grounded* in the situation and *oriented* to something. From the navigation point of view, attention can be seen as a mental control: it focuses perception on something as an object of situational relevance. What is relevant for the person depends on his/her life situation at the time of perception. Background information is in diachronic relation to the ongoing situation: it provides the referential background. Past and future are in such consideration like two hemispheres opening from the person's perspective[44]. Yet, they differ in the sense that the past is a *unique* trajectory of individual experiences, whereas the future involves a *choice* (alternative paths).

[43] Koffka (1935): figure/ground in perception

[44] In many languages, the past has been symbolised and conceptualised as something in the back, whereas future opens in the front.

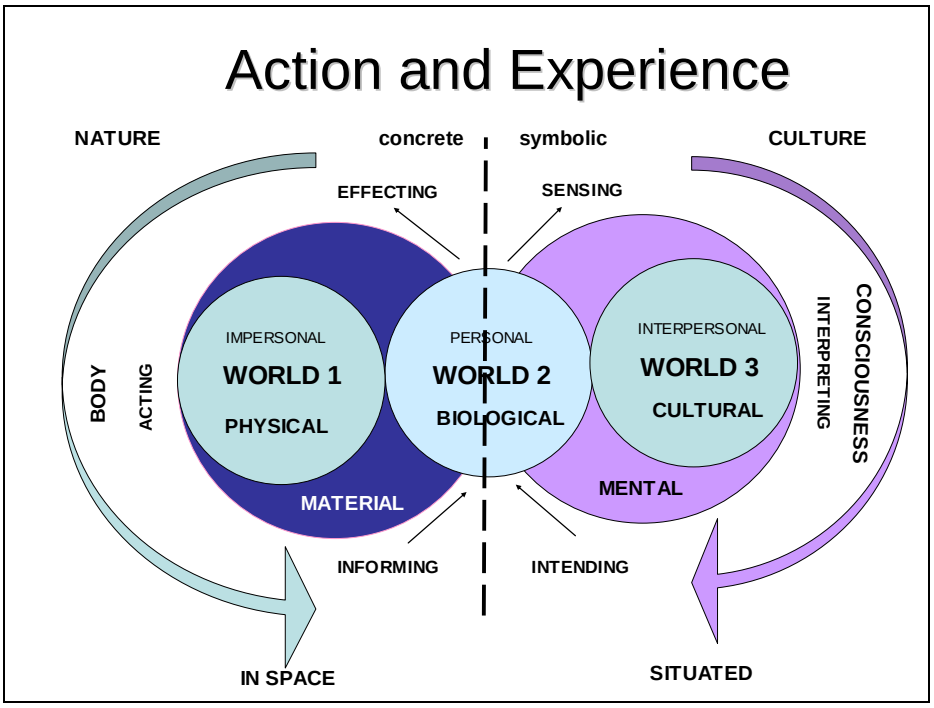


Figure 5 Human experience and intentional action
Human body is the interface of the person's internal and external world. The body is the necessary 'instrument' for subjective experience and for 'executing' intentional actions (through motor actions) in the spatial behaviour.
Human body is also the instrument of communicative acts through verbal (speech) and nonverbal (e.g. visual-gestural) modalities. Language and different social conventions are learned (> internalised) in the cultural context where the person has been brought up/ lives.

Making a distinction between human and non-human objects is one of the precondition for human communication; yet, recognizing the object as human object is not yet sufficient: understanding the *intentions* of others while watching their actions is fundamental for social behaviour (Iacoboni & al, 2005). Therefore, the object has to be classified as another human object, that is, the origo of another human perspective / subjective experience.

Diessel (2006) points out joint attention as foundational for communication. The attention of the addressee has to be captured in order to initiate a communication. Furthermore, the latter has to respond to the invitation by acceptance in order to activate the connection across the two perspectives. Focusing the gaze on a person and/or calling him/her verbally are ways to establish *mutual* attention. Mutuality activates the channel between the communicating parts, making it possible for them to refer to a common object, and to establish joint attention on it (>a referential triangle) (Figure 6).

When a small child is concerned, the focus of his/her gaze provides the spatial coordinates of the object of his/her *interest.* A communicative situation between the adults is, however, more complex: people may reveal their focus of interest by their spatial behaviour, but they are equally capable of hiding their motivaton for strategic reasons. To take an example from the Western culture, mutual visual attention is expected when people shake hands. People may signal for courtesy fake interest on an object. Also, when people reflect and gather their thoughts, they tend to look into distance or fix their gaze on a random object as if taking space for mental concentration. In co-action, people tend to be in their 'engagement space' aware of the focus of the co-participant's visual attention (Gill, date not available).

In human communication, the question is not about *reactions* to something but about *responses* to one another. Attention is the mechanism to trigger mutuality. Yet, mutuality is conditional, because to be successful, attention has to be responded to by the addressee. Attention distribution can then be manipulated on objects of interest. Pointing is one of the techniques to guide visual attention to something in the shared information space. The one who wishes to introduce an object to the attention of the addressee, indicates in the direction of the object. When the addressee's attention turns to the object, a referential triangle between the communicating actors and the object of reference is established. It then provides a common ground for further contributions and their coordination.

To summarise different views of attention from a workplace design perspective, attention can be regarded as the connection of a person and the object of his/her reference/relevance. *The scope, the orientation,* and *the focus* are structural characteristics of attention. Focusing attention on a human object has *a communicative function* (for instance opening of a mutual communication). Attention manipulation is also expected to follow expectations of context-appropriate behaviour (=social norms). Visual attention in a communicative situation is detected/determined from the person's gaze direction/focus.

Both verbal and nonverbal modes of communication are used to guide the addressee's attention. Within a reach of touch, haptic signalling can be used. In order to evoke intended response, the signal has to be *salient* enough to have impact on the addressee's behaviour. In terms of quantity, excessive intensity and duration of attention tend to have negative impact on communication. In terms of quality, empathetic attention evokes (affect) responses whereas indifference rather smoulders them.

Joint attention on an object is a referential triangle that connects the positions of the participants and the position of an object that they focus on in their *shared* information space. It provides a *joint situational perceptual referent* (Figure 6).

3.4 Deixis

Deixis is the act of reference, context-boundedness; the object - for instance an item - is not only what it is, and where it is, but also a *meaning* attributed (assigned) to it (e.g. Peirce, 1868). The position of a material object is a *physical* aspect of the reference, whereas meaning is its *intentional* aspect. Individual meaning has a *situated* and *triadic* nature: it is something appearing as something to someone in a particular situational context.

Shared information space is a consensual domain, a common ground for references in a communicative situation. A triadic structure connects two individuals' perspectives on a common object in their shared information space.

Usage-based theory of language acquisition recognizes *mutual attention* and *joint attention* as precursors of verbal communication (Diessel, 2006). Focusing attention on one another establishes a relationship between a dyad. Joint attention to an object is a more advanced achievement implying a *theory of mind* (Tomasello, 1999), that is, an understanding that another person has her own egocentric perspective onto the world, structurally similar to one's own but content-wise different.

Cognitive linguistics shows that *declarative pointing* is a particular form of communication employed in human communication (Diessel, 2006). Other primates do not present objects to their conspecifics' attention by pointing, and even human beings have to achieve it in their early caretaker – child interactions. Mutual eye-contact connects two perspectives and indicates their mutual attention. In a baby's experience, being in the centre of human attention is, at the same, to be cared[45]. The children become gradually able[46] to follow the gaze of the caretaker, first on an object that is within their common visual frame, and later on, they start to follow the gaze beyond it, turning to look in the direction of the caretaker's attention. Before they learn to speak, children rely on gestural communication; for instance, they indicate by pointing to an object of their interest/desire.

[45] Note here the affect-laden aspect of an eye-contact.
[46] Note also Piaget's (1950) studies on the child's constuction of reality.

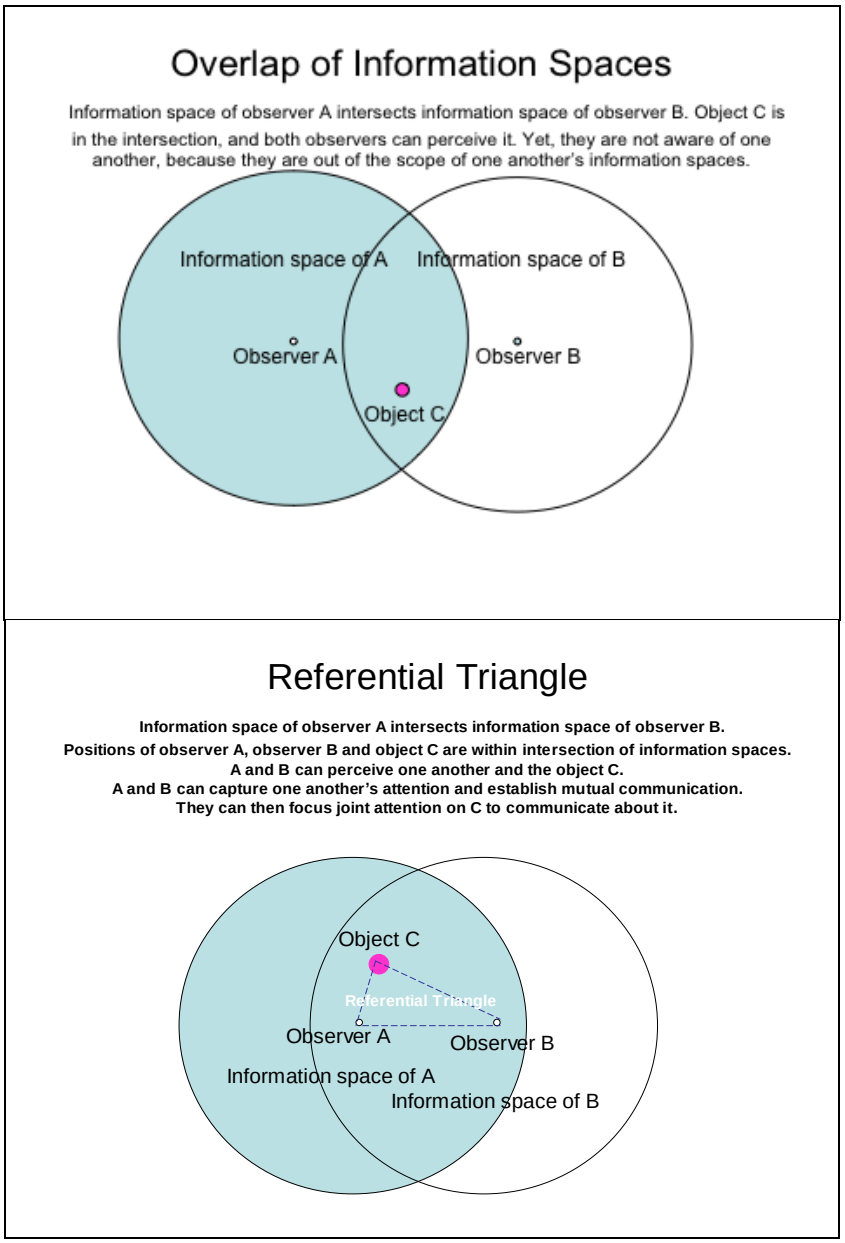


Figure 6 Communicative situation and a joint attention
People have to be within one another's perceptual field in order to communicate with one another: one has to catch the attention of the other in order to initiate a contact, and the other has to accept the invitation in order to activate their mutual communication.
Both participants can perceive a material object if and only if it is within the intersection of their perceptual fields (joint attention as a referential triangle in the bottom picture).

In a workplace design consideration, deixis is a *triadic structure* (=subject/object/meaning) that is anchored through the situation to the physical space. It describes individual perspective and an individual meaning. A(nother) triadic structure describes a *communicative situation*: it connects the position of the person who refers to an object, the position of the person who is addressed in a communicative situation (Figure 6, lower

frame), and the position of the object indicated in their shared information space[47]. Its communicative function is to provide a *joint situational referent* in communication.

Hand pointing is a *deictic gesture*[48]; its *communicative function* is to guide the addressee's attention to focus on a particular object (referent). From the social point of view, it is a tool in the interpersonal navigation; it provides cues for the addressee to make assumptions of the intentions of the speaker. Therefore, it contributes to the formation of shared understanding through

- the focus shift in the space (direction, spatial orientation),
- the shift within/out of a topical frame (state of communication), and
- the shift in the cognitive state of the participant (background knowledge / motivation / intention).

The interaction space that is designed for collaboration should make it possible for the participants *to see what they need to see, to hear what they need to hear, and to do what they need to do in order to carry out a joint task*. In other words, it should enable *mutual orientation* and *joint orientation*, in a topic relevant scope and in (a) relevant mode(s).

3.5 Common Ground in Conversation

Deictic practices are an integral part of human communication. Mutual *responsiveness*, *commitment* to the joint activity, and commitment to *mutual support* are the three foundational features of a joint effort (Bratman, 1992, 327-328; Nardi, 2005). Each participant has individual commitment to it. The participants seek to take into account the intentions and actions of one another, and to guide their behaviours in a responsive way (Bratman, 1992, 340-341). The participants also support one another in carrying out their roles in the joint activity (ibid, 328).

Conversation is considered the fundamental site of language use. Cognitive linguistics considers conversation as *a collaborative process*, where the speakers and the addressees contribute in co-operative way *to establish a definitive referent* (e.g. Clark & Wilkes-Gibbs, 1986; Clark & Shaefer, 1989; Clark & Brennan, 1991; Clark, 1996; Clark, 2001). The unit of conversation is called a *contribution*.

Clark and his colleagues explain the complementary role of speaking and understanding in the coordination of the contributions. The speaker and the addressee(s) play the

[47] E.g. Hutchins (1995) takes a distributed view on cognition, and Järvilehto (2000) regards consciousness as co-operation.
[48] Note gesture studies by McNeill (2005; 2008) and Kendon (e.g. 1967 and 1990).

principal roles of a conversation: the person in voice seeks to make sure that s/he is attended to, heard, and understood. The content of speech is moulded not only by the speaker but also by the addressee: the speaker's utterance is by foundation designed in relation to whom the utterance is intended to be (e.g. Clark & Krych, 2004).

A new contribution to a conversation begins by presentation phase; in the acceptance phase, the participants work together to establish a mutual belief that everybody has understood it in order to move on in the conversation (Clark & Shaeffer, 1983, 19-41; Clark & Krych, 2004). Not only is timing critical for the coordination of the contributions[49] but also the contents of what other contributors have presented by the time of a new entry: it may be an illocutionary act[50] such as a promise to the addressee, it may be a request to do something, or, to provide additional information.

Clark explains *common ground* at the language level as *background knowledge / mutual beliefs / mutual assumptions assumed by the participants by the time of a new entry*. People infer their common ground from the past conversations, joint perceptual experiences and joint membership in cultural communities (communities of practice) (Figure 7). They have to coordinate on three parallel things: the content, the participants, and the roles, in order to maintain their common ground in the course of a conversation (Clark & Schaefer, 1989; Clark, 1996).

Clark focuses particularly on the language level. In such consideration, common grounding is stratification of sequential propositions. Common grounding in a conversation extends, however, far beyond the linguistic domain. The so-called back channel plays an important role: the participants tend to inform in nonverbal or paralinguistic ways whether they have gathered what the speaker communicated. Social conventions provide a common frame for a context-appropriate turn taking and language use. People have learned to 'read' and interpret other participants' social (gestural) cues, and to exploit them when designing their entries and synchronizing them with the flow of conversation. Skilled participants[51] (Gill & Borchers, 2004) know *how* and *when* to respond appropriately for the purpose at hand. The participant's own mental context model (v Dijk, 2006) is therefore involved in the design of a new entry. Timing and manner are behavioural aspects in the coordination of contributions.

[49] According to Détienne (2006), co-operative processes relevant to collaboration include coordination to manage task dependencies, establishment of common ground, and negotiation mechanism to manage the integration of multiple perspectives.
[50] Austin's speech act theory (1964) includes illocutionary acts.
[51] Note here v Dijk's (2006) mental context model.

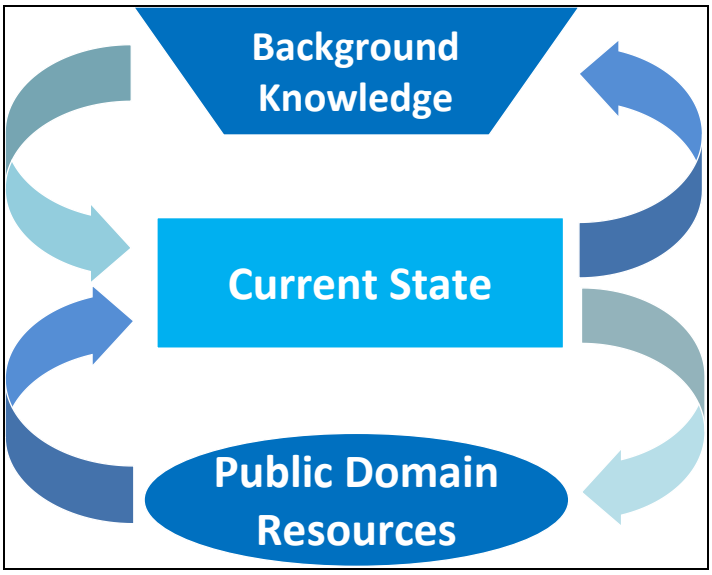


Figure 7 Common ground, background knowledge, and public domain resources
(from Rosenberg & al. (in public domain 2003), People-place-process framework in SANE-Human Environment Modelling). People update their common ground in a conversation (=current state) drawing on their situational background knowledge and informational resources that are publicly available / are brought to their perceptual field (public domain resources).

From the workplace design perspective, the speaker's utterance is not only sound output from a particular source (>perspective) but it involves an intention (>direction): each participants of a conversation has an individual spatial *perspective* wherefrom they navigate the physical, social and conceptual space. Language is a tool for guiding the attention of another person to objects of one's reference. Both visual and audio information play an important role in language use: a conversation unfolds step by step when people coordinate their perspectives and moves as a sequence of contributions to a common topic (>a joint frame), and at the same, update their individual/common background knowledge (>current state of a conversation).

An analysis of the navigation strategies (that are applied in language use) should therefore include the following:

- Co-occurrence of a pointing gesture and a verbal reference
- Use of interrogatives in reference to objects in the spatial settings
- Use of demonstratives in connection with pointing behaviour / subsequent impact on visual attention of the addressees;
- Pointing gesture / subsequent affirmation / negation by an addressee.

Shared information space is the *functional bridge* between participating perspectives. Visual connection is important in the procedural coordination of contributions (speaker/addressee role switch)[52]. Verbal communication requires, apart from mutuality and reciprocity of perspectives, a *common topic* as a frame of co-operative action. Within such 'overlap frame', new objects of reference can be introduced in *verbal* and *nonverbal* means.

3.6 Cross-Modality in Communication

The studies of human communication, child language acquisition, and developmental cognition recognize that people communicate through verbal and non-verbal channels, and that the two modes of communication complement one another (Darwin, 1872). In addition, non-verbal communication is a modal precursor of verbal communication. Gesture is considered a critical link between human conceptualizing capacities and linguistic abilities (McNeill & al, in Jaimes, 2007). Gestures are typically employed when speakers have difficulties finding words, or when they try to articulate new ideas into verbal expressions (e.g. Roth, 2000; 2002; McNeill, 2005). They can be used more or less successfully as a replacement of verbal utterance in a situation where conventional speech communication is not possible.

Though linguistic resources are regarded as crucial to the construction and reconstruction of social order, also other modalities are involved. Roth (2004) argues that *perceptual gestalts* should be treated in the analysis of communication at the same level as other *semiotic resources* drawn on in communicative interactions. The unit of analysis for (pragmatics) studies of communicative action should therefore account for three modalities: *verbal*, *gestural* and *perceptual modalities* (ibid).

Perceptual gestalts are semiotic resources that the participants of collaboration take as 'something that goes without saying' (ibid.); in other words, they are available to and known by other participants of that situation. Communication during collaborative work relies much on perceptual gestalts. Perceptual gestalts also have a role in turn taking.

In a social encounter, multiple modes of communication are used in parallel and in complementary ways, flexibly switching the emphasis from one channel to another depending on the situational context. For instance, while speaking, people use properties of voice to convey subtle meanings; they gesticulate, may change a posture and position,

[52] Note here also gestural communication in the guidance of the attention of the audience.

or carry out other types of communicative action, such as placing an object somewhere for someone (e.g. Kendon, 1990; Roth, 2004; Clark, 2005).

Extra-linguistic acts (such as a nod) are functional equivalents of the properties of talk: parallel functional gestures prop up states of talk and police them (Goffman, 1964): a person may, for instance, cue through a posture change and redirection of attention that s/he is leaving. Also intonation contour towards the end of an utterance tends to cue the next move. Extra-linguistic acts are therefore an important source of procedural information.

Gestures are used to direct attention of relevant communicators, disambiguate information, and express size, direction, movement, and attitude. They provide a common ground for people who do not share the same language. Robertson (2000) notes that pointing *embodies action,* and their relevant affordances: it is about something emitting signs, monitoring of signs, moving in and out of a shared space. From the communication point of view, highlighting some aspect of an object by dividing the domain of scrutiny into a figure and its background makes something 'stand out'. Highlighting embodies not only one's perception but serves to direct the attention of others.

Affective reasoning plays a role in social interaction: *affect displays* (Darwin, 1872) provide information of the situational state of a co-participant. According to McQuiggan & al. (2007), people assess one another's situational context, modify their own affective state accordingly, and respond to these outcomes by expressing empathetic behaviours.

In computer-mediated communication, a number of ways have been developed to provide contextual cues and mitigate adverse effects of technology-mediation (e.g. telepointing[53]: Isaacs & Tang, 1994; eye-contact: Vertegaal & al, 2003; gaze awareness: Gemmel & al, 2000; weak gaze awareness: Ohno, 2005; augmented reality: Billinghurst & al, 2002; Barakonyi & al, 2004).

From the workplace design perspective, cross-modality is a particularly challenging issue in the design of distributed/virtual settings. Visual information conveys social cues that are important to the coordination of individual contributions. Therefore, the individual perspectives have to be spatially 'orchestrated' in a way that provides a *congruent perceptual 'common ground'* for the participants' collaboration.

[53] The effect of a telepointer on performance and preference has been studied e.g. by Adams et al. (2005).

3.7 Affordance in Design Discourse

Affordance discourse is reviewed in the following starting from Gibson's ecological perceptual definition, and providing in comparison to it a selection of views presented in HCI-related literature.

In Gibson's definition, affordance has a dual nature; it cuts across the subject-object dichotomy, being simultaneously "a fact of the environment and a fact of behaviour" (Gibson, 1979, 129). Even though there is an 'instrumental' interface in the human eye (or any sensory organ), Gibson considers the perceiving individual and her environment as something inseparable. In that view, perception is direct and immediate, without significant intermediate states involving memory or inferences. Affordances of the environment are

> *"what it offers the animal, what it provides and furnishes, either for good or ill"* (Gibson, 1979, 129).

Affordance is *action potential* latent in the environment; objectively it is independent of the actor's ability to recognize it but *actually* dependent on the properties of the actor.

Norman – real affordance / perceived affordance: Norman approached affordance from an information processing point of view: he first considered Gibson's construct perceived affordance that is, as laden with past knowledge and experiences (McGrenere & Ho, 2000, 3). Later on (1999), he made a distinction between 'real affordance' and 'perceived affordance'; affordances are properties of the world, and they reflect the possible relationships among actors and objects.

In practical design, usability issues, such as critical clues for operation in a local case of use, have to be considered. 'Perceived affordance' is therefore an important design concept. For instance the computer system comes with in-built, real affordances; the designer can control perceived affordances in the graphical interfaces. The three main concepts to understand how to operate a new device are the underlying conceptual model, constrains, and affordances. (Norman, 1999, 38-42).

A coherent, understandable design comes through an explicit, perceivable conceptual model. Norman classifies three different types of human behavioural constraints: physical, logical and cultural constraints[54]. Physical constraints set the limits of the scope of possibilities. Logical constraints use reasoning in deducing alternatives. They are

[54] Note: Yet, Norman does not mention psychological constraints as prejudice, inhibition, and phobia.

indispensable for a good conceptual model that has to have an intelligent fit to human cognition.

Norman also points out the difference between *conventions* and *affordance*. Conventions are cultural, human made constraints that prefer particular activities to others. Cultural interpretations of symbols have an important role in design. Cultural conventions always imply a community of practice; they are shared among a wider group of people, and while being collective in nature, they tend to be slow to change. Conventions have limited strength because they can be ignored or violated; nevertheless, they provide an important tool in the navigation of anything unknown.

Gaver - perceptible, hidden, false, sequential, and nested affordances: Gaver (1991, 79-83; 1992; 1996) was among the early contributors to the affordance discourse in HCI-community who addressed technological affordances.

For Gaver, affordance provides an integrated account of complex configuration attributes to address a broad range of interface issues. Factors of perception and action determine whether or not designed interfaces are easy to learn and use. The notion points to a configuration of properties, implying physical attributes of the things to be acted upon that are compatible with those of the actor, that information about those attributes is available in a form compatible with a perceptual system of the actor, and additionally, that these attributes and the action they make possible, are relevant to culture and a perceiver (ibid.).

Gaver provides a framework for separating affordances from perceptual information available about them (ibid.): perceptual information may suggest affordances that do not actually exist; on the other hand, there can be affordances that are difficult to perceive. Making affordances perceptible is therefore a way to design easy-to-use systems: perceptible affordances offer a direct link between perception and action.

Perceptible affordances are inter-referential as the attributes of the object relevant for action are available for perception. Yet, the actual perception of affordances is partly determined by the culture, social setting, experience and intentions of the observer (ibid.).

The notion of affordance implies exploration and learning: affordances are explored, and the exploration of afforded actions leads to discovery of the system. In that regard, learning is a matter of attention[55]- rather than inference. Acting on a perceptible affordance leads to information indicating new affordances. Affordances unfolding in

[55] Through attentional focus learning receives a spatial direction grounded in the spatial settings of action!

time are in Gaver's classification 'sequential affordances', and affordances grouped in space are 'nested affordances'. Complex actions can be understood in terms of groups of affordances that are sequential in time, or nested in space, and in terms of the abilities of different media to reveal them (Gaver, 1991).

Furthermore, Gaver (1996) points out affordances that emerge due to interaction and actions in the environment such as communities of practice and social conventions.

McGrenere & Ho - affordance as design framework for usability: McGrenere & Ho (2000) provide a critical discussion of the concept within HCI-discourse by the year 2000, and expand affordance into a framework for usability design. They define the fundamental properties of an affordance as existing relative to the action capabilities of a particular actor, independent of the actor's ability to perceive it, and independent of the changes of the needs and goals of the actor.

In that meaning, affordance is an idea of *actor-environment mutuality*: an affordance is invariant, and its existence does not depend on interpretation; yet, an affordance is dependent of a particular actor as a frame of reference.

Designing the *utility* of an object and designing the way in which that utility is conveyed are two different questions: the latter is of *usability* of the object, the design of the perceptual information specifying affordances. The easiness to use, usability, also varies from user to user. For instance, the time taken by a novice user to undertake affordances is typically different from that of an expert user (ibid. 3).

McGrenere and Ho (2000) point out that usability of a design can be enhanced by clearly designing the perceptual information that specifies these affordances. In addition to being consistent, providing appropriate feedback and error recovery, usable designs have information specifying affordances accounting for attributes of the end-users, their cultural conventions and level of expertise. In software design, affordances are functions invocable by the user of the software. Yet, affordances do not necessarily map one-to-one on system functions because they may be nested in a hierarchy: for instance in GUI-interface the editing menu nests different functional options. The design models have a *metaphoric* nature, and the metaphor applied gives the system a particular set of affordances. The design of forms can convey non-visual information (in the form of sound, taste, smell, and texture). For instance the size and shape of input devices that have high level of freedom should be designed to be manipulated by fingers in order to provide more accurate control than does an arm control.

The goal of the design, in McGrenere's and Ho's view, is to determine the necessary affordances, and secondly, to maximise each of these dimensions. The *quantitative* aspect of affordance is therefore focal in the evaluation of design, because thinking in terms of either-or affordances does not lead anywhere in assessing degrees. To put in quantitative terms, usability has fuzzy nature: it is about *more-or-less*. The basis of evaluation lies in the relation of the amount of perceptual information, and easiness to undertake an affordance: when the time to perform the action is reduced, an affordance is easier to undertake. The degree of an affordance exists *relative* to a particular user, and the goal of usability design is to achieve the optimal point for the target user.

Hartson - cognitive, physical, sensory and functional affordances: Hartson (2003, 315-338) contributes to make affordance applicable as a concept for interaction design and evaluation. Drawing on Gibson, Norman, and a few other authors, his formulation builds on semantics: it includes four types of affordances, each relevant to a different type of user action[56] relevant to interaction design. User action is abstracted to different types of action in order to differentiate between pure sensing, perception, and cognition.

Each type of affordance plays in his consideration a different role, has different mechanisms, corresponds to different kinds of user actions, exhibits different characteristics, has different requirements for design, and implies different things in evaluation and diagnosis; therefore each type of affordance must be identified what it is for, and considered on its own terms in analysis and design.

- A cognitive affordance[57] in Harson's classification is a design feature that supports or enables thinking or knowing about something. Design of cognitive affordances is about learnability and easiness-to-use[58].
- A physical affordance[59] is a design feature that physically helps or enables doing something. The design of physical affordances is design of usability in the form of high performance and productivity for a multiple type of users.
- Sensory affordances help users with sensory actions, such as salient colour to bring figure 'out' of the background and thus support its perception.
- A functional affordance is the fourth type of affordance, by which Hartson refers to the link between usage and usefulness. He uses the term 'functional

[56] In this, Hartson builds on Norman's stages-of-action model, 1986.
[57] Hartson calls Norman's 'perceived affordance' in his classification 'cognitive affordance' because it helps users with their cognitive actions.
[58] About cognitive part of usability
[59] 'Real affordances' (in Norman' classification) are called 'physical affordances' in Hartson's classification.

affordance' to denote higher-level user enablement in the work domain. For instance a doorknob (physical affordance) can be grasped and turned opening thereby a door. It is used to operate the door that enables passage, its *functional affordance*. A physical affordance gives only access to that functionality.

Usability stems from

> *"…the effectiveness of cognitive affordances for understanding how to use physical affordances, from the physical ease of using the physical affordances, and from the sensing of these via sensory affordances*". (ibid, 321-322).

Sensory affordances play, in Hartson's view, a critical supporting role in interaction design: the users must be able to sense cognitive affordances and physical affordances in order for them to aid the users' cognitive and physical actions. Therefore, in the design of user interface artefacts noticeability, discernibility, legibility and audibility is important.

Hartson (2003) points out that the ontology of affordance is relational: the existence of affordances is relative to the environment of the users and usage. Effectiveness of an affordance depends on the both the use and the artefact and their attributes. The user must sense, understand, and use affordances within an interaction design in order to accomplish work goals. Each affordance role is involved in both learning about and using artefacts: physical affordance is associated with operability; cognitive affordance with semantics (meaning of the user interface artefacts), and sensory affordance is associated with sense-ability of the user interface artefacts and its characteristics.

Turner – simple and complex affordances; affordance as context: Turner (2005) makes a distinction between simple and complex affordances. The former refers to basic level ergonomic aspects of usability, whereas the latter denotes purposes of activity in a cultural context[60].

Turner equals affordances to context. A more accurate way would be, to my understanding, to equate affordance to the human-environment interface as a structure that has potential for action (capacity). That is the way Hartson (2003) ends up with his description of affordance[61].

[60] Applying the concept to computer supported collaborative work, Turner also uses a middle layer of affordances, referring to user tasks and their embodiment. (For more see Turner & Turner, 2003).

[61] Hartson's and Turner's definitions of affordance appear congruent with existential phenomenology and emergent materialism: it is an interface resulting from the physical and biological evolution, and from the social-historical development. That is much alike the view I took in an unpublished conference paper at InterSymp 1996 Conference in Baden-Baden, Germany, and later on in my Licentiate thesis (in Finnish) at Helsinki University of Technology, Finland, 2000: I articulated the human-environment relationship into

To conclude affordance discourse from a workplace design point of view, it describes the human-environment interface as a multifaceted and multi-level structure. Hartson's (2003) articulation that makes distinction between the physical, sensory, functional, and cognitive aspect of the interface, and Turner's (2005) interpretation that helps to articulate its social (cultural) aspect, are particular informative in the design of interaction space.

3.8 Summary and Conclusion

The chapter presented notions relevant to an investigation of deictic practices and to an articulation of the human-environment interface (Table 2).

Affordance is a notion that refers to a *human-environment interface as a multi-level/multimodal structure/action potential*. In the light of multidisciplinary theory in human communication (Chapters 2 & 3), the human-environment interface can be described as follows:

> *Human being is an intentional actor that has metric extensions and sensor-motor capacities in a life situation where s/he is a unique perspective onto the world in the spatial settings within a particular social-cultural context; in it, she participates in communication in a (complementary) role having an individual history of experiences (trajectory), and thereby, particular knowledge background as a frame of interpretation when s/he orients, motivated by an individual value system (based on affects and reasoning), to something in his/her perceptual field (situational information space) paying attention to it as an object of reference/ relevance/interest.*

My conclusion of the reviewed literature in the role of space in human communication (Chapters 2 & 3) is the following:

Human cognition is, from a workplace design point of view, an individual perspective within the settings of collaboration; it is an origo of attention/navigation, and a frame of interpretation based on individual background knowledge in reference to situational information, and (in intentional orientation) to the future. The position of a person embeds (situates) an individual perspective to the settings as a multisensory (>multimodal) interface. Communicative situation arranges individual perspectives in a spatial relation.

---

different layers resulting from physical, biological, cultural and - what I called - 'egological' processes of evolution. Such articulation parallels emergent materialism and existential phenomenology as complementary approaches from the opposite directions.

Table 2 Key concepts in the investigation of human communication from a workplace design perspective

Notion	Notion described from a workplace design perspective
Communicative situation	• Locus of communication as spatial distribution of components involved at a particular time (=time slice). • The coordination site that gathers individual contributions. • A distribution site for individual contributions / a site for the participants to draw on situational information.
Experience	• Referential (>intentional) perspective of a sensor-motor unit in its spatial surrounding. • The coordination site of personal motivation and orientation.
Attention	• Mental control of orientation. • Signal mechanism that has orientational and coordination impact on co-operation. • Modal connection between the participant and the object of observation/reference.
Deixis	• Act of reference in the spatial / speech context. • Spatial deixis 'grounds' the content of communication through the speech context in the communicative situation (>settings & individual perspectives).
Common ground	• The state of shared understanding in a conversation at the time of talk.
Cross-modality	• Parallel/complementary use of different modes of communication, mode switches included.
Affordance	• The overall structure/capacity of the human-environment interface for action.

(See references in Chapter 3.1-3.7).

The methodological insight from developmental and cognitive psychology is that joint visual attention and verbal common ground share similar pragmatic function in the coordination of joint action: pointing (as a communicative gesture commonly used in face-to-face communication) both embodies intentional action, and specifies the position of the object of reference in the spatial surroundings.

From a *workplace design* point of view, an integrated approach is required in the investigation of communication/ collaboration: verbal utterance has to be taken into account in concert with gestural communication in order to capture in the analysis pragmatic and semantic aspects that are relevant to the coordination of individual perspectives through the spatial settings.

An appropriate unit of analysis is *a communicative situation*. Thereby, communication can be analysed as a sequence of time slices (that is, changes of a communicative situation in the course of communicative acts/events), and interpersonal interfaces can be focused on in a *spatial*, *modal*, *social*, and *semantic* consideration.

In order for a framework to guide the design of effective interaction space, it has to cover not only *the levels* relevant to the coordination of individual perspectives but also *the modal aspects* of *a human-environment interface*.

The following Chapter 4 describes a multidisciplinary framework that was developed for an investigation of deixis in interpersonal communication with the aim to guide the design of functional affordances for shared understanding (which is a condition of a coherent joint action) - independent of whether or not the settings are collocated. (The framework was refined through the investigation of deictic gesture in the video-mediated settings (Chapter 5)).

4 Framework for Investigating Deixis in Interpersonal Communication

The chapter presents an analytic framework for investigating deixis in the interpersonal communication (in particular, the problem of a deictic reference in the video-mediated communication). It provides a multidisciplinary explanation of shared understanding, and deixis-based design principles for workplace design.

The framework considers a human actor *in a communicative situation* in terms of

1. *human-environment interface* (>individual multisensory/multimodal perspective grounded by foundation in the physical space (>subjective experience))
2. *human-human interface* (>mutuality through a shared information space)
3. *intentional action* manifest in the person's *spatial behaviour.*

The explanation of deixis is achieved through situating the human-environment interface *and* the human-human interface in a communicative situation as a conceptual structure. The framework bridges a number of deixis relevant arguments from different schools of thought[62], by locating them in the framework according to their scope of validity.

In a physical consideration, human communication is, at the bottom, (signal transmission and) information processing[63]. In a phenomenological consideration, spatial behaviour and mental interpretation are two *different aspects of the same situation*. Though mental contents are not reducible to physical terms, they *receive spatial coordinates* through the person's corporeality in a life situation where individual meaning making (=signification process) takes place[64] (e.g. Malpas, 2002; spatial cognition: Tversky, op.cit; Gibson,

[62] Among them are holism of context (Malpas, 2002), inseparability of organism/environment (e.g. Gibson, 1979); situation as a site of a social encounter (Goffman, 1964), ecological view of visual perception (Gibson, 1979), figure/ground articulation in perception (Koffka, 1935), perceptual gestalt in communication (Roth, 2004), triadic relation of sign-object-interpretant (Peirce, 1868), framing of experience (Goffman, 1974), origo of subjective orientation (Buehler, 1934), spatial cognition (Tversky, 2003), situated cognition (e.g. Roth, 2001), perspectival systems in spatial navigation (Tversky, 2005), reciprocity of perspectives and joint referential basis in intersubjectivity (Schutz, 1973), shared intentionality (Tomasello, 2009), joint attention in interpersonal communication (Diessel, 2006), socially situated cognition (Smith & Semin, 2007), spatial patterns of focused social encounters (Kendon, 1990), spatial patterns of social practices (Alexander, 1977, Hillier & Hanson, 1984), cross-modality in interpersonal communication (Roth, 2004), common ground in language use (Clark, 1996), context model in episodic memory (van Dijk, 2006), anthropocentric design approach (Vitruvius).

[63] It implies a temporal dimension: in information processing theory, a meaningful unit of information is *chunk*, and a basic unit of behaviour is *test-operate-test-exit circle* (Miller, 1960). Note also Shannon's (1948) mathematical theory of communication.

[64] Sign basically stands in a triadic relation: it is to someone for something in some respect. Sign refers to object through likeness (>icon), through actual connection (>index), or, through convention (>symbol) (Peirce, 1868).

1979; Buehler, 1934).

In a modal consideration, people communicate through multimodal *deictic practices* (Roth, 2004; Diessel, 2006). Individual views are expressed in *communicative acts* that are executed in *spatial behaviour* in a *social-spatial situation* (Goffman, 1964). Visual-gestural domain plays a prevalent role in the *procedural coordination* of individual perspectives: it serves *the establishment of mutuality* in a shared information space, and in a speech context, it serves *the guidance of attention*, and thereby, *the coordination of the individual perspectives on a common topic/object*.

In a social consideration, human beings are born and bred within social practices (Gabora, 2008); therefore, a social context not only *surrounds* them (as more or less institutionalised conventions (e.g. Norman, 1999), and 'inbuilt' social logic in the built environments (Hillier & Hanson, 1984)) but *lies within* themselves.

In a semantic consideration, communication is a continuous *interpretation* and *learning process*[65] within a cultural context: when people communicate, they *interpret the situation* on the basis of their state-of-art knowledge, beliefs, and assumptions that they infer from prior experiences (Clark, 1996, 332; mental context model: v Dijk, 2006; framing: Goffman, 1974). Though social conventions provide predictability in social situations, there is also a less predictable element in the form of misunderstanding and (un)intentional breach of context-appropriate behaviour[66].

In a methodological consideration, the framework (presented in Chapter 4.1) accommodates the cognitive approach and the ethnographic approach without blurring them. From a practical design point of view, it accommodates reciprocity of perspectives *and* multi-/cross-modality of human communication in a design relevant way.

4.1 Multidisciplinary Explanation of Deixis in Interpersonal Communication

A communicative situation[67] covers, to put it in interrogatives, the *who*, *what*, *where*, *when*, *what for* and *how* of a collaborative effort. In this framework, a communicative situation is called as a structure LOCUS. It is a spatial temporal node that coordinates the positions of the participants of communication in their spatial settings, their ICT-

[65] Dewey's constructivist view on 'learning by doing'; *experiential learning*: Kolb, & Fry (1975).
[66] Note also that creative process is somehow stepping out of the conventional frame.
[67] See Goffman (1964) for social encounter unfolding in the physical settings*, and* Kendon (1967, 1990) for interpersonal encounters as spatial performance (gaze direction, mutual positioning and orientation).

connections included. Through a spatial distribution, a communicative situation determines the participants' relative positions at a particular time.

Individual perspective. A participant of a communication is a spatial body in a particular position; the position thereby functions as an interface that determines what situational information from the surrounding world is available to/for a participant. Moreover, it coordinates that situational information in a cross-modal way: for instance, when orienting in their surroundings, people rely on what they can see, hear, feel, and even what they can smell[68]. The position is at the same the origo of the participant's motor actions including speech, gestures and a manipulation of objects that can be observed in the surroundings. In this framework, the position of a participant is called MODUS as a node that coordinates the situational information and the participant's actions[69].

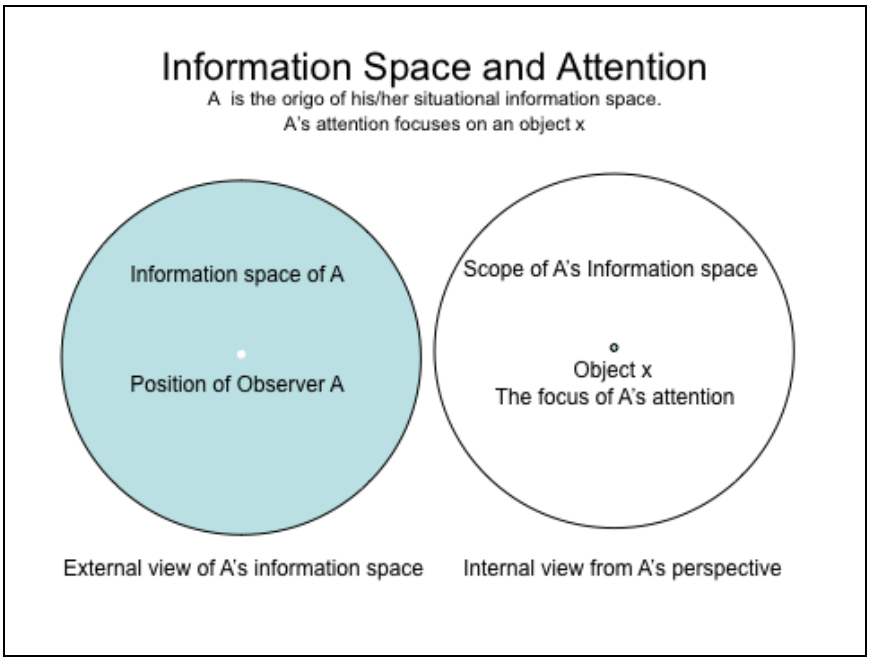


Figure 8 Complementary views on individual information space
Each person has in his/her individual information space. It can be illustrated as a scope of situational information (= information available for the person's sensory caption and subjective experience at a particular time). A situational perceptual field of a person (A) can be described in metric terms through the distance (radius) and the direction (scope/sector) from the position of A. In observation, the person's attention focuses on something (object x) in his/her perceptual field (=situational information space).

[68] Compare with *verbal, gestural* and *perceptual* modalities in communication/collaboration (Roth, 2004).

[69] Compare with ecological psychological view of *inseparability* of human/environment (Gibson, 1979 > affordance) and with Tversky's (2008) *spatial cognition* as *situated* and *embodied*.

Apart from nesting the participant's body as a spatial unit, the position also nests his/her cognitive and motivational state; moreover, the trajectory of his/her individual experiences[70]. In such consideration, the position functions as a situational origo of subjective orientation/motivation in the world[71].

Attention[72] is a mental control operating at the human-environment interface; it orients the mind in a particular direction within the scope of situational information (Figure 8). In such regard, attention is a navigation tool in the service of signification. To take an example, a particular item in the surroundings is attended to as information that is registered through sensory modalities, articulated in perception into a figure out of its background[73], and identified on the basis of a situational worldview (=background knowledge) *as* something[74]. The outcome, that is, a subjective meaning attached to the object is a situational intentional object in reference to a situational 'real' object.

In the line of spatial/informational encounters and their subjective experiences, attention functions as a node of coordination and translation between situational information and the participant's background knowledge and motivation.

The fact that a participant pays attention to an object indicates a situational relevance between an attention and a motivation: there is a reason *why, what for* particular object appears to a participant as *'inter-esting'* in a particular situation. To take an example, there is a sudden sound; the participant registers it due to its salience, and turns in its direction to find out what it was about - whether it was possibly a sign of danger that would motivate (further actions in) response to it. The fact that the participant registers the sound is a *functional* aspect of attention, whereas his/her seeking to figure out what it was about is an *intentional* aspect of it (covering affect and reason). In a motivational consideration, attention functions as a self-regulation tool in the service of adaptation to a context change.

In this framework, FOCUS is a node of coordination that describes a participant's attention in a communicative situation through his/her position *and* the position of an object attended to. If attention is paid to a material object located in the surrounding

---

[70] Compare with Rauhala's (1993) existential-phenomenological view of experience as *a semantic layer* in human existence, with ecological psychological views (Gibson, 1979), and with studies of *situated* and *embodied cognition* (e.g. Merleau-Ponty, 1945, Tversky, 2008).

[71] Compare with Buehler's (1923) *Ich-Hier-Jetzt* as origo of subjective orientation

[72] *Attention* as the most basic function of human communication (Diessel, 2006)

[73] Compare with *figure/ground* demarcation in perception (Koffka, 1935)

[74] Compare with Peirce's (1868) triad *sign-object-interpretant*.

space, the origo and the focus of the *visual attention* of a particular participant can be defined in the *spatial* coordinates. Thereby, also the participant's situational subjective motivation and the situational meaning attached by him/her to a situational object receive spatial coordinates. Moreover, the situational visual perceptual field of the participant can be mapped through distance (radius) and direction (scope) as a situational *visual* information space.

In communication, people discriminate between *human* and *non-human* objects[75]. In this framework, we call such an understanding internalized reciprocity of perspectives: though another person's perspective is as a conceptual structure similar to one's own perspective, its origo is in a different position and consequently, it provides a different view onto a shared information space; moreover, it nests a unique individual worldview based on an individual trajectory of spatial encounters, and thereby, a different history of experiences and their interpretations.

Though a participant's perspective in a communicative situation can be defined as a spatial structure, more requirements have to be met in order to explain a (possibility of) a shared understanding for workplace design. The first condition in order for an interpersonal communication to be coherent is that the individual perspectives (in space) and contributions (in time) have to be coordinated.

Coordination of the individual perspectives. While there is no shortcut between individual minds, the individual views have to be expressed in order to inform co-participants of a communication. In order for communication to flow between the individual perspectives, there has to be an information channel between their positions. The channel availability and reliability constrain the *mode* of communication: when at a speech distance, people use both the verbal and the nonverbal channel; should there be a glass wall between them, they would have to rely on nonverbal communication, or, to use, in addition, a communication tool that enables verbal communication.

Compared with the collocated settings, ICT-mediation poses particular constraints to communication in terms of the mode, (the scope) of the shared information space, and technical quality of the connection: it would be useless to communicate through hand pointing on the phone, or, to carry on communicating to a remote person after a signal failure has broken down the connection. In an ordinary videoconference again, a joint

[75] See studies on early baby-caretaker interactions, in particular the role of mutual eye-contact, gaze following and joint attention (e.g. Tomasello & Carpenter, 2007), note also *shared intentionality* (Tomasello, 2009).

visual frame is too challenging for the human cognition to infer the referent of an across-site pointing. Even though the remote site participants can *see* the act of pointing, they are not able to identify the object of reference; thereby, a shared understanding fails - *unless* the pointing person provides cross-modal information in reference to the object.

In order to construct something coherent in a collaborative way from their individual perspectives, the participants of collaboration have to have at least some motivation in common to establish a joint frame, that is, *what* they are going to do together. Is it perhaps a project set and fixed in advance, or, is it something to be negotiated? A common frame is, to a degree, given by the cultural background through learning[76]: for instance, whether the participants speak the same language and whether they know one another's cultural background. In this framework, a joint framing is called, as a node of coordination, TOPIC: it constrains the scope (=what), the motivation (=what for), the channel/modality (=how), and the role allocation (between whom) in a communicative situation[77].

A call for attention and its acceptance are a condition for reciprocity of perspectives to be established, and for a mutual communication to take place. A communication has a beginning and an end. Depending on the situation, attention can be called for through the verbal and/or the nonverbal channel; for instance, focusing a gaze at a person who is at a close distance is a powerful way to draw attention. 'Hi there!' is a sufficient call to catch the attention of a person at a speech distance, and tapping on a shoulder works within the reach of a hand. Similarly, the acceptance/declination of a call - as well as closing mutual communication - is signalled through the verbal and/or the nonverbal channel.

Role allocation and *turn taking* in the complementary roles serve the coordination of individual contributions into a joint line of communication[78]. In a dialogue for instance, the complementary roles are the speaker and the addressee (=listener). A contribution to a conversation is a subjective point of view, expressed in an active speaker role through the verbal and/or nonverbal channel in reference to the topic, and to (a) prior contribution(s). Co-participants' contributions are attended to in the role of a listener/spectator. Turns are taken in the complementary roles, and a turn taking is signalled in different ways depending on the situation: through intonation, gestures, and/or verbal expressions. In

[76] See e.g. Gabora (2008) for *social cognition* and *socially situated cognition,* Kendon's (1990) *pattern* in social encounters, Norman (1999) for *convention* and Clark's (1996) for *common ground* in language use.

[77] See Bratman (1992) for foundational features of a joint effort: *mutual responsiveness, commitment to joint activity* and to *mutual support*); *reciprocity of perspectives* as foundational for socially organised interactions (Schutz & Luckmann, 1973).

[78] See contribution theory of conversation (Clark, 1996).

many social-cultural practices, the role allocation has consolidated also into particular spatial patterns[79], to take the actor – the audience interface in a theatre as an example.

Apart from *mutual* attention (to establish a reciprocity of perspectives), a collaborative effort requires a *joint* attention[80] to enable a shared understanding. Shared understanding is defined in this framework as an assumption that the other participants of communication identify the situational object referred to by the contributing participant. Shared understanding is a condition for a coherent line of communication where it serves as a *symbolic common ground* for further contributions.

Shared attention is in a spatial respect a referential triangle that connects the position of the speaker, the position of the addressee, and the position of the object referred to (Figure 9). The participant who is introducing a new object has to guide the co-participants' attention to the object of reference: therefore, s/he has to call the attention of his/her audience - for instance through a gaze contact - and indicate the location of the object (should it be a visual object in their spatial surroundings), by turning to look in its direction, maybe also pointing to it.

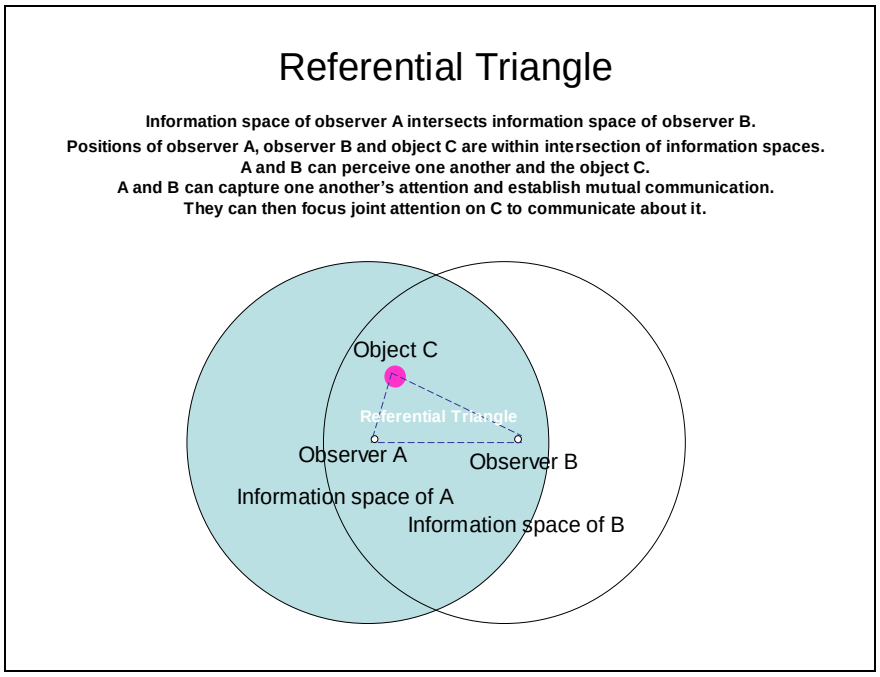


Figure 9 Referential triangle
The participants can both perceive a tacit object if it is within the intersection of their perceptual fields: if person A points to an object that is within A: s and B:s shared information space, person B can establish a joint attention with A to the object (C).

[79] See *patterns* in social encounters (Kendon, 1990) and *social logic of space* (Hillier & Hanson, 1984).

[80] See Diessel (2006) for the role of *joint attention* in communication.

In addition, s/he has to monitor the *impact* of his/her pointing and/or verbal expression in its *communicative function* (=guiding attention). If the other participants respond by turning to look in the same direction, it is taken by the contributing participant as a positive sign: s/he gathers that the participants have identified the object of his/her reference because they have established a joint visual attention to it; s/her *assumes* that they have got a shared understanding of it[81], and his/her attention is free to move on to the following object.

Though a topic of a communication is more or less fixed, its focus moves along with the individual, mutual, and joint attention when new objects of reference are introduced.

Apart from the *functional* conditions, a possibility of a shared understanding is dependent on the *cultural* conditions[82]. To take an example, language use is a cultural phenomenon: in a functional consideration, it relies on sound but in order for the participants to understand a verbal message, they must have *learned* the same language. Yet, even then, misunderstandings occur because each participant is an individual 'meaning maker'. Misunderstanding is a sign of a failure in reaching a shared understanding. As there are no two identical views from two different angles and from different participants' cognitive perspectives, neither can there be any shared understanding of an object of reference with a 100 % certainty. In an interpersonal communication, however, there is no other way available than a trial-and-error: therefore, an individual *assumption* of a shared understanding is taken as a 'good enough' common ground to build on – until its foundations might, in the context of further topic relevant information, turn out to be too wavering to hold. Once discovered, a false assumption is revised and repaired through negotiation, that is, by requesting and providing additional information in reference to the object, before moving on to the following object.

4.2 Design principles for Collaborative Settings

The practical purpose of this framework is to inform design in the analysis of the *functional* and *cognitive ergonomic* requirements of collaboration (>> usability criteria), and in the design of *functional affordances*[83] for collaboration.

The research question is *how to provide functional affordances for shared understanding through the spatial settings.*

---

81 Compare with Clark's (1996) *common ground* in language use.

82 See *symbol theory* by Elias (1990) and *complex affordance* by Turner (2005).

83 According to Hartson's (2003) articulation of affordances

In order for the framework to inform design, the design relevant aspects of deictic practices have to be translated into spatial/temporal (>> structural/functional) terms. The underpinnings of shared understanding are described in this framework as a multi-layer structure of an interpersonal interface (Figure 10). In it, the participant's position provides *a spatial perspective* to the object, and the participant's experience/background knowledge/motivation/scenarios provide *a cognitive (> diachronic) perspective* to it.

Shared understanding can best be explained for design through *visual modality*; this is because joint visual attention on an object can be defined as a referential triangle between the position of the participant who is referring to an object (by looking / pointing to it), the position of the participant who is observing the action, and the position of the object referred to. (Though communication modes differ in terms of radius and scope, each of them is in principle a field opening from the person's position).

Due to the complexity of deixis in interpersonal communication and collaboration, several types of phenomena and their inter-relations have to be taken into account. First, human communication is multi-/cross-modal *intentional action* in its nature, and second, each sensory modality has specific *functional conditions*.

ICT-mediation further complicates the picture because the distributed settings differ from the collocated settings as an *operational (modal) connection* and as a *frame of interpretation*: a technical channel constrains the *scope* and the *sensory modality* of a shared information space and thereby, the possibility of shared understanding. In addition, (the origo and the frame of) a camera view differs from the participant's point of which complicates mutual orientation.

Yet, whether a natural or an ICT-mediated situation, the connection is, at the bottom, a *physical*, *informational* connection. *Interaction space* is therefore, from a workplace design point of view, *a modal connection that enables mutuality and joint attention on the topical objects in the task relevant mode(s).*

In this framework, the *point of departure* is set on the human-human interface (Figure 10). The context of a deictic reference (the context of interpersonal meaning making) is described as a modal (=sensory-operational) connection between the participants' (A and B) perspectives. A temporal change (e.g. opening, closing, movement, construction) is regarded as a transformation of a communicative situation. In practical design, the point of departure is therefore in *complementary (role) perspectives*.

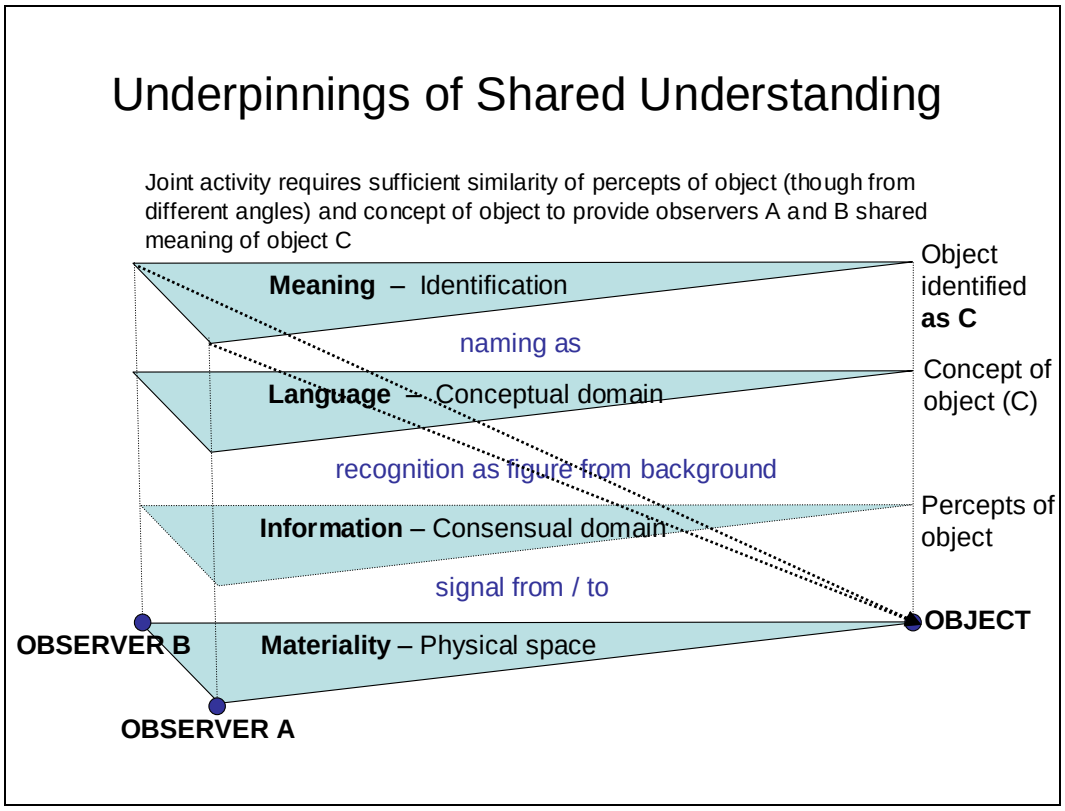


Figure 10 Shared understanding as a multi-layer structure

The framework describes A's and B's shared understanding of an object (referred to by A or B) as an outcome of their deictic practices (=verbal and visual gestural communication).

1. As a structure, both individual perspectives are grounded in the physical space.
2. In sensory perception, a figure is created in each participant's perspective from the informational resources ('background') at the time of perception. (Compare with Figure 7.)
3. A figure becomes interpreted in a situational communicative/cognitive context (=background knowledge in a communicative situation). Verbal referencing occurs at language level.
4. The outcome of interpretation is an individual meaning of an object of reference, that is, the object as something for someone in particular situation.

A and B (assume that they) have a shared understanding of the object (C).

An example: The speaker A informs the addressee B in a verbal way by uttering something while turning at the same to look at an object in their physical surroundings. The addressee B gathers the gestural message through visual information, and the verbal message through audio information. B identifies the object of A's visual attention in their shared information space (that provides a congruent spatial frame for their orientation). Though B focuses on the same spatial object establishing *joint attention* to it, s/he perceives it from a different angle, and understands it in an individual way. In order for the A and B to create a shared understanding of the speaker's verbal utterance in reference to the object, they have to share a common frame at the level of language (that is, they have to speak the same language). Only then they are able to negotiate individual meanings, and establish a shared understanding (C) of what A wanted to communicate in reference to the object. From the workplace design point of view, shared understanding of an object of reference is *an outcome of a multi-layer coordination of individual perspectives/meanings*. Four nodes of coordination are foundational for the facilitation of shared understanding: locus, modus, topic and focus (Table 4).

A possibility of a joint attention is limited to the objects *within the intersection of the individual perceptual fields* (=the participants' interaction space) (Figure 9). The participant's perspectives have to be interconnected in an effective way in order to guide the addressee's attention. Therefore, the perceptual/operational fields (> relative positions of the participants) have to be coordinated in a design in a way that provides a *multi-perspective* to *a shared information space* (=a joint frame) in which objects can be introduced to the attention of co-participants.

Due to the complexity of human communication, the perspectives of the participants have to be coordinated through four orders to enable (a joint attention on an object, and thereby, a possibility of) shared understanding. The nodes of coordination are called here LOCUS, MODUS, TOPIC and FOCUS (Table 3; Table 4).

- Who/where? - Locus refers to the immediate surroundings of the communicating participants (actors).
- How? - Modus refers to the reciprocity of perspectives as an intersection of individual perceptual and operational fields. Human-computer interface and technology-mediated representations are regarded here as an artificial extension/constraint of a natural connection.
- What for? - Topic refers to a shared referential frame. Translated into spatial design terms, it is the scope of a shared information space due to cover the topic relevant objects in the task relevant mode(s).
- What/where? - Focus refers to an individual / mutual / joint attention to an object. In spatial design terms, focus moves along with an attention in the shared information space (where new objects of reference are brought by the contributing participant to the attention of co-participants in a task relevant mode).

The *topic* of collaboration constrains (through a role allocation) what needs to seen / heard / manipulated (from particular role perspective) in order to contribute, and to carry out topical tasks; a design question is therefore *what kind of modal connection is required* to address (a range of) topical objects from complementary perspectives.

In a design, a relevant set of controls (e.g. scope > frame; direction > focus) has to be provided in order for the role perspectives to guide attention in an effective way. A focal object requires *a relevant and informative context*; a technical way to guide attention to an object of situational relevance/reference is through the relevant framing and adequate focusing (of a camera view, sound input/output).

Table 3 Designing enabling settings for shared understanding

PARTICIPANTS as	**COORDINATION OF PERSPECTIVES** to enable collaboration		**SPATIAL SETTING as**
SPATIAL BEINGS 'taking space/ place'	**SPATIAL DISTRIBUTION OF PARTICIPANTS**	1. **'LOCUS'** See 'physical space' in Figure 10	**PARTICIPANT'S (ROLE) POSITIONS** in the layout
SENSORY(-MOTOR) BEINGS registering information from their surroundings	**MUTUAL CONNECTION** = **INFORMATION CHANNEL** between **PARTICIPANTS' PERSPECTIVES**	2. **'MODUS'** See 'consensual domain' in Figure 10	**SHARED INFORMATION SPACE** >mutual awareness
SOCIAL BEINGS using language, gesture =multimodal communication	**COORDINATION OF ROLE PERSPECTIVES** in relation to objects to be discussed / manipulated **> CHANNEL REQUIREMENTS** in terms of modalities	3. **'TOPIC'** See 'language level' in Figure 10	**INTERACTION SPACE > FUNCTIONAL AFFORDANCES** for participants' collaborative tasks
COGNITIVE BEINGS having each their own intentions	**COORDINATION OF CONTRIBUTIONS** through turn taking	**5. 'FOCUS'** See 'meaning' in Figure 10	**CONTROLS** for guiding co-participants' attention to object of reference

A participant of communication is considered in workplace design in four different orders:

1. a spatial being collocated or distributed with other similar beings (=co-participants);
2. 'a moving set of sensors' (as Tversky et al. (2008) put it) registering information;
3. a social being capable of *symbolic* communication, and
4. a cognitive being having personal intentions to be communicated through language use and gestures (=bridged via the physical sphere as there is no shortcut between two persons' minds). Turn taking and guiding attention are methods of coordinating contributions. Joint attention to an object implies *referential triangle* (See Figure 9).

4.3 Summary

The context of a deictic reference (in other words, interpersonal meaning making in a collaborative situation) was described in a way that is relevant to workplace design. The explanation of deixis can be summarized as follows: an *information channel* between the participants' positions is the *communicative bridge* between their spatial (>>informational) perspectives. What people can do jointly in a situation depends on the *scope* and the *control* of their *interaction space*; in other words, it depends on the *functional (modal, operational) capacity of their connection to convey task sensitive information in (a) topic relevant mode(s) between the participants' spatial perspectives.*

The framework presented in this chapter (Figure 10, Table 3, Table 4) guides the design of functional affordances for shared understanding. It is named *the locus-modus-topic-focus framework*[84], because four coordination nodes are foundational in the facilitation of (a definitive referent, and thereby,) shared understanding in collaboration. Interaction space is, in such consideration, a modal connection of the participants(' positions) that enables mutuality and joint attention on topical objects in (a) task relevant mode(s). The following Chapter 5 describes the two case studies that were conducted to test the viability of the framework. The design principles (for the facilitation of shared understanding through the settings) can be summarized as follows:

Provide

- an information channel (& relevant controls to the complementary role perspectives) in the topic relevant modes between the positions of the participants, in order to enable
- a shared information space, to enable
- mutual awareness, to enable
- mutual attention, to enable
- a joint attention on task relevant objects *in the shared information space* (=interaction space), to enable
- a shared understanding of an object of reference, to enable
- a coherent joint line of contributions (=effective communication/ collaboration).

This deixis-based approach represents a human-centred design tradition. Compared with other approaches, there is a difference in the method and in the coverage:

- It *situates the human-human interface* in *the communicative situation*, and thereby, it encompasses also the modal and social aspects of communication that are relevant to the coordination of perspectives *(>reciprocity of perspectives)*, and to the guidance of attention to the object of reference.
- It considers *ICT-mediation a functional extension/constraint of the communicating participant's natural information space*; thereby, it encompasses the participants' *perspectival in/congruency*.

It coordinates the participants' perspectives *through the settings* in a way that provides *an interaction space* that is both topic relevant and task sensitive to their collaborative tasks.

[84] Alternatively, HHI-framework because it describes the human-human interface.

Table 4 HHI-Framework: locus, modus, topic and focus

Individual perspectives are coordinated within a stack of constraining frames.	**Node of coordination**	**Spatial design issues at different coordination levels / TECHNICAL ASPECT**
Joint attentional focus in conversation / co-operation, through reciprocate manipulation of attention	**FOCUS**	State of a conversation (co-operation as a joint orientation to a situational object) Situational focus of joint attention
		FLOOR CONTROL (>relevant focusing and framing of camera view; control of sound input/output)
Topical 'common ground' as a joint referential frame (=what people are doing in a collaborative way) Congruent conversational / co-operative frame	**^ TOPIC**	Scope of communication Scope of collaboration Cultural frame
		FRAME OF SHARED INFORMATION SPACE (>interaction space)
Perceptual 'common ground' as informational reciprocity of perspectives (Joint referential frame)	**^ MODUS**	Features of the connection
		FEATURES (>two-way audio-visual connection, pointing functionality)
Spatial distribution of objects Spatial 'common ground' (Co-existential frame)	**^ LOCUS**	Spatial configuration
		INFORMATION CHANNEL

In design, the participants' perspectives are coordinated through four different organisational nodes (=locus, modus, topic, focus) to provide functional affordances for shared understanding: what the participants can do jointly in a situation depends on the scope and the control of their interaction space (=the functional capacity of the connection to convey task sensitive information (= direct FOCUS) in a topic relevant mode (=MODUS, TOPIC) between the participants' perspectives (=LOCUS/FOCUS).

LOCUS: the position of a participant at the time of communication is where an individual meaning is coordinated. Therefore, it is a situational spatial frame.
MODUS: The settings have to provide a two-way connection between the participants' positions in the mode(s) relevant to/employed in communication. (Verbal and visual-gestural modes of communication are characteristic of social situations > audio/visual channel). The object of reference has to be (or, has to be brought/requested) in the shared information space (=consensual domain) in order for the participants to perceive it. Shared information space a situational informational frame.
TOPIC: In order to communicate, the participants have to have some motivation to it, and in order to negotiate meaning, they have to have on a cultural level a joint frame (>language they both understand). Topic is therefore a joint frame at the cultural level.
FOCUS: In order for the participants to identify objects of reference, they have to guide one another's attention to the object they refer to in their shared information space. In such guidance, verbal and visual-gestural communication modes are typically employed (>audio/visual channel). Joint attention on an object is therefore a coordination site as a joint situational frame of reference. It is foundational for the establishment of a shared understanding of the object of reference: it provides the situational bridge between the individual participants' perspectives, and its changes coordinate a coherent line of communication. Joint attention is therefore a coordination site as a situational orientational frame.

5 Methods and Analysis of Data

The chapter describes the procedure and the selection of the methods (Figure 12, Table 5, Table 6, Table 8)[85] used to validate the proposed framework (Chapter 4) in the two case studies. To start with, *deictic gesture in the video-mediated communication* is presented as a practical problem from a spatial design perspective.

Case study 1 focuses on explaining communicative aspects of gestural deixis in the video-mediated settings. Learning from its outcome, the design instructions derived from the framework are then applied to a practical design case (Case study 2).

As concluded in Chapter 2.7 (pp.36-38), the cognitive and the ethnographic approach are two complementary approaches to interpersonal communication and collaboration: the former focuses on *the individual actor's perspective* whereas the latter has emphasis on the social-spatial patterns, *communicative practices in their cultural context*.

In this study, empirical and experimental methods are integrated in a compatible way to explain the mechanism of gestural deixis for workplace design. Experimental methods are commonly used in the studies of perception and attention. In Cognitive Linguistics, for instance, they have been employed to explore referring as a collaborative process (Clark & Wilkes-Gibbs, 1986). When the aim is to describe communicative practices as they unfold in real-life situations - such as work practices in an organisation -, different methods are relevant: collecting data through stakeholder interviews and focus group discussions elucidates different role perspectives and their (inter)actions as described by the actors themselves, whereas ethnographic observation provides an external view that, at the same, seeks to understand observed events as if 'being there', participant of that culture (e.g. Saville-Troike & Trudgill, 1982; Suchman, 1986; Simonsen & Kensing, 1997, Rosenberg, 2003).

When products and systems are developed, their practical usability is evaluated in user tests (e.g. Norman, 1988; Nielsen, 1993; Nielsen 1995). The approach taken in Case study 1 is qualitative due to the communicative function of a pointing gesture in video-mediated communication: the usability of the experimental setting is evaluated by analysing video episodes where the test users are carrying out collaborative tasks in the setting (Table 6). A complementary technique is analysing narrative data from a post-test questionnaire in order to learn from user experience for further stages of design.

[85] Related documents are available to the official examiners of the thesis.

5.1 Investigating Communicative Function of Pointing Gesture in Video-Mediated Settings

If you point to an object, people nearby can locate it without any difficulty. If you point in a videoconference to an object located at the remote site, people there have no idea which of the objects in their surroundings you are pointing to - unless you describe it (Figure 11). This is why gestural deixis is a particular challenge for spatial workplace design, in particular for hybrid[86] interaction design.

Deictic reference is still an under-researched topic from the practical workplace design point of view: though the role of deictic gesture in human communication has been described in the multidisciplinary literature (e.g. Kendon, 1867, Mc Neill, 2005; 2008; Diessel, 2006; Enfield et al, 2007), the *communicative-functional* basis for inferring shared understanding of the reference object has not been explicated to inform workplace design in the *hybrid* interaction.

Therefore, this research seeks to find out whether it is possible to enable the communicative function of an across-site pointing gesture in the video-mediated communication (Figure 11). The focus is on the following questions:

- What properties of interaction space are crucial for inferring shared understanding of the object of a pointing gesture?
- What do we need to know to inform spatial design in effective gestural deixis?

Figure 11 Research question: gestural deixis in remote communication[87]

[86] Hybrid' here refers to synchronous communication between people some of whom are collocated and others non-collocated.

[87] Video frames above are from the tests run on Test bed 1.

5.2 Progress of Study

The research commenced by investigating the role of space in a communicative situation. Aggregating empirical data through ethnographic observation and video recording[88], the focus was initially on the identification of the patterns of communication in the collocated vs. video-mediated settings (Figure 12, Figure 14, Table 5).

In the investigation, a dual approach was taken: a communicative situation was considered

1. spatial distribution of material components (the participants included)
2. common informational / referential basis of
    - the participants' cognitive states, and
    - the participants' communicative actions.

The findings from an observation how people use space *in the collocated* communication (Table 5) provided basis for investigating gestural deixis *in the video-mediated settings*. For such purpose, an exploratory method was employed in Case study 1: a hypothetical assumption of video-mediated pointing was constructed from the observational insights in combination with theoretical views on deixis, and a test scenario was drafted (Chapter 5.4). A design instruction was then outlined for the implementation of a test bed to investigate whether the underlying assumption (of the communicative function of an across-site pointing gesture in the video-mediated communication) held ground. Furthermore, the settings and the test tasks were designed (Chapter 5.4), including the consent request and the instructions for the participants. Apart from hard data from video recording to establish *effectiveness* of the settings in enabling the communicative function of an across-site pointing gesture, complementary narrative data of *user experience* was collected through a short questionnaire and feedback discussion after each test session (Appendix 5, Appendix 6).

The method for analysing data was developed (Chapter 5.4.3) focusing on the spatial and modal (>verbal, nonverbal) aspects of a communicative act, with respect to its impact on co-participants' communicative acts (Figure 12, Figure 14).

[88] Consent of the participants for video recording.

Table 5 Method of initial data collection before specifying the research questions

METHOD	THREE APPROACHES COMBINED to compare communication in the collocated vs. video-mediated settings:		
	Architecture	**Human Communication**	**Human –Computer Interaction**
ETHNOGRAPHIC OBSERVATION focusing on	-Use of space in communication /collaboration: How space is used in a collocated / video-mediated situation	-Interplay of verbal and nonverbal modes of communication -Role of gestures (gaze direction, pointing gestures) in concert with speech -Patterns of interpersonal communication -Spatial -Procedural	-Differences between collocated (natural) and ICT-mediated communication: -Deictic reference in collocated vs. video-mediated settings
VIDEO RECORDING to investigate	-Communicative situation as a shared frame for orientation/action -Communicative situation as the common referential basis for communication	-Situations where pointing gestures occur -Focus of attention -Joint attention -Guidance of attention -Referential triangle Success/ failure/problem in establishing understanding of an object of reference	-Success/failure/ difficulty in establishing shared understanding of an object of reference in collocated vs. video-mediated settings
AIM	**TO FOCUS & TO SPECIFY RESEARCH QUESTIONS**		

In Case study 2, the design principles[89] derived for the settings for hybrid interaction were guiding a practical design case: a lecture theatre due to be converted for video-mediated lecturing. The point of departure was then in

1. activity as a spatial/social performance and
2. user perspective (=> teacher, student, AV staff).

The settings were implemented according to the principles that combine (the spatial and social-cognitive aspects of)

1. mutuality as *a functional connection* between individual perspectives, and
2. deixis as *a referential triangle* (Figure 9).

The settings were assessed in a showcase session focusing on

[89] Derived from the framework, taking into account the findings from the 1st case study.

1. technical robustness (=>effective connection), and
2. user acceptance (perception of the settings as a frame of action / interpretation).

Feedback from the initial tests was fed forward to refine the design (Figure 13).

Data was eventually collected also of initial use, including the users' ways to avoid and overcome problems. Practice-based information was assumed to support emerging practice, and to foster the community of practice to share their experiences for mutual learning – not only for the benefit of the novice users but also for horizontal learning across different stakeholder perspectives (teacher / student / AV staff).

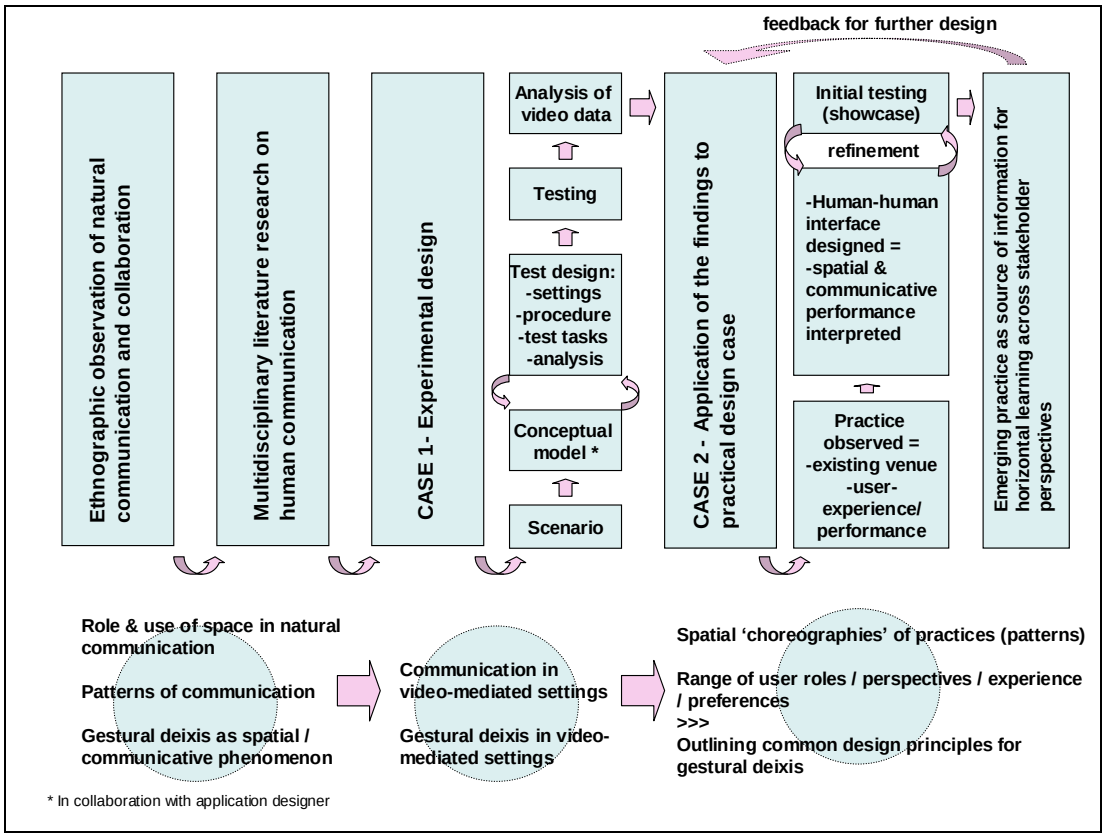


Figure 12 The progress of the research

5.3 Methods Applied

The following methods were employed to address the questions outlined in Chapter 5.1:

- ethnographically informed methods (observation and video analysis) for collecting empirical data;

- exploratory method (including the implementation of the test-bed) to investigate the communicative function of pointing gesture in the video-mediated settings, and to derive design principles for effective gestural deixis in the hybrid interaction[90] (to guide the design of the settings for video-mediated collaboration).

5.3.1 Ethnographic Observation

Ethnography is a qualitative method, originally developed in Cultural Anthropology to identify and describe patterns such as rituals[91]: The researchers used to spend long periods of time in the investigated culture to gain understanding of them in the way they make sense to the members of that specific culture. Later on, ethnographically informed methods were adopted for discipline-specific purposes beyond their initial field of application: for instance, in Communication studies (Savill-Troike & Drudgill, 1982), Human-Computer Interaction, and Computer Supported Cooperative Work (e.g. Suchman, 1987; Robertson, 2000).

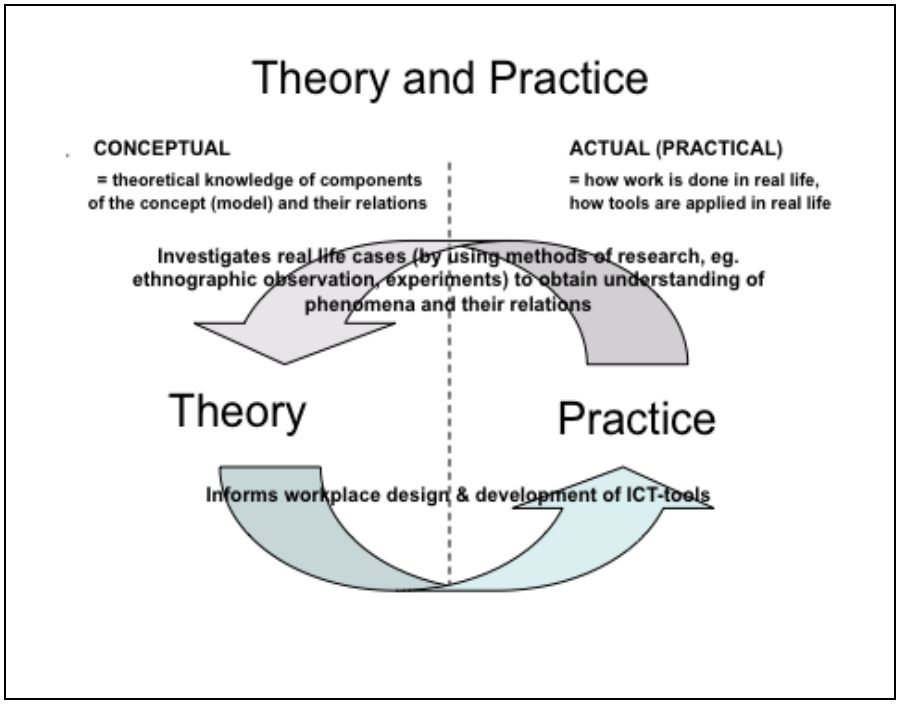


Figure 13 The relation between theory and practice in the field of design

[90] 'Hybrid' communication here refers to synchronous communication between people some of whom are collocated and others non-collocated.
[91] E.g. coming of age in Samoa (Mead, 1928). Note also Mauss' (1924) studies of reciprocity and exchange.

Ethnographic observation was applied in the investigation as it appears to be an effective method for discerning communicative patterns and to identifying the role of space in communication (Figure 14). Yet, the risk of subjective bias, a typical pitfall of participant observation, has to be kept in mind. Using parallel methods provides a firmer basis for interpretations and conclusions. Among them, reviewing the findings of an observation with the informant helps to decrease at least some of the risk.

5.3.2 Video-ethnography

Video-ethnography has become common in the studies of human communication; it has been applied for instance to the doctor-patient interaction in the surgery (Heath & Hindmarsh, 2002), and to staff members' ad-hoc encounters in the hospital ward corridor (Iedema & al., 2005). The strength of the method is in providing pictorial data for a detailed analysis of communicative phenomena: through video data,[92] communication and collaboration can be analysed as *a sequence of communicative acts* and *their impacts.*

Video data[93] captures seating position, gaze direction, facial expression, and hand gesture as a spatial composition at the time of their occurrence. It provides pictorial evidence to support or call to question interpretations drawn from observation - which is indispensable in the investigation of gestural deixis. It has particular strength in discerning the role and the interplay of different communication modes.

In this research, video-ethnography helps to analyse, frame-by-frame, events, co-occurrences, trajectories and temporal relations, and analyse them both from the spatial and communicative point of view. It is then possible to explore nonverbal communicative acts, in particular gaze direction and pointing gesture, in relation to parallel talk and deictic reference (Figure 20).

Snapshots and video frames can be used to discern behavioural strategies that people use to trigger their addressees' responses, in particular: in guiding their attention to an object of reference (Figure 31).

As we have no direct access to subjective experience of another person, shared understanding has to be inferred from observable phenomena. Assumptions of the communicative participants' shared understanding have therefore to be pieced together

[92] Video-ethnography has been described in a somewhat different way, e.g. in Heath, C., & Hindmarsh, J. (2002).

[93] Ethical issues were formally discussed in a meeting with the Director of PhD studies and the supervisor of the thesis. Consent of the participants was requested for taking photos/videos and using frames for illustration of the findings.

from circumstantial evidence such as the addressee's gaze turning in the direction of the indicated object (Figure 22), confirmative (non)verbal expressions and successful co-operative actions.

In this investigation, pictorial data (snapshots and video clips) was used to identify the gaze direction of a participant in relation to another participant's actions in order to locate the situational focus of (gestural) communication (=>an object of reference) (Figure 14).

Time	Gaze	Hand gestures	Position / attitude	Talk
5.35.35		A bringing forth a business card.	F continues leaning, B leans forward to the table.	Continues
5.39.48	B looking in his front.	A turning to hand it over to B who is ready to take it.	F continues leaning, all smiling.	Continues
5.41.55	B looking at A, C at B.	B holding the card with both hands.	F leaning, B and C smiling a broad smile.	Continues
5.42.02	B studying the card, A follows by sight.		F leaning, all smiling.	Continues
5.43.28	B, still studying it, A watching her, C looking at A.		F leaning, all smiling.	Continues
---	---	---	----	---
6.01.33	B and C looking at A who is looking in his front.		Everybody straightens precisely at the same time. A, B and C smile.	Continues
6.02.33	A looking at people across the table, others looking at him.	C takes a grip of the edge of the table.	All laugh (as if of relief). All in the position of taking off.	Ends in mutual thanks.
6.02.99	Everybody looking in their front.		People take support when standing up, all still smiling.	
6.03.46			The whole group stands up at the same time.	

Figure 14 A sample of natural observation in a collocated situation
Ethnographic observation and video-ethnography were used to gain understanding of how people use space (=multimodal deixis in a collocated situation). The above clip reveals the subtle role that nonverbal cues play in concert with verbal utterances in face-to-face communication: towards the end of the session, a long series of nonverbal social cues prepare the group for a conclusion. When a context-appropriate point (=thanks, smiles) has been reached, everybody stands up within a fraction of a second and leaves.

5.3.3 Exploratory Method

Experimental test set-up is a scientific method for testing hypotheses and validating a concept/theory. Hypotheses can be based on a theoretical construction, insights drawn from empirical data, or, both of them.

Usability testing is a particular type of experimental study commonly applied in the development of new products and systems (e.g. Nielsen, 1993). In usability tests, 'asking nature' is not enough: quantitative methods apply to the investigation of functional aspects, but when *user perspective* has to be taken into account, qualitative methods play an important role.

In Case study 1, exploratory method was developed to investigate interaction space that is required for successful gestural deixis (=>effective gestural communication) in the distributed collaboration. Test settings were designed for the test users' communicative action and interpretation. A scenario of an effective across-site pointing was based on the theoretical views of deixis, combined with the findings from the ethnographic observation and video data of collocated natural communication.

The test settings (videoconferencing system with pointing functionality (Figure 15)) were designed in a close collaboration of the application engineer and the researcher (Appendix 2) to facilitate across-site pointing and thereby, to make it possible *to explore the impact of a communication tool on deictic practices.*

The test tasks (Appendix 4) were due to provoke communicative behaviour, and to yield data for assessing whether or not it is possible to extrapolate the communicative function of a pointing gesture to video-mediated collaboration.

Figure 15 Test settings
In the settings for one-way pointing (left), the camera was above the right side display. In the settings for two-way pointing, the camera was positioned between the two displays.

Scheduling the tests and finding convenient time for the test users was not a simple matter. It was challenging to design test tasks taking into consideration that people are seldom willing to offer their time for a long test; yet, the tasks had to be covering enough to clarify the impact of a communication tool on collaborative efforts. At the same, they had to provoke a range of co-operative behaviours.

5.4 Case Study 1: From Collocated Pointing to Remote Pointing

The first case study explored in the experimental settings *the communicative function of a deictic gesture in the video-mediated communication* in order to determine whether it is possible to mediate *the communicative function of pointing gesture* across different locations. It also aimed to find out whether the participants manage to adapt to the artificially augmented frame of action/interpretation, and whether or not they started using pointing functionality in their communication.

Case study 1 was based on the following assumption:

- A joint attention is a condition of shared understanding.
- Shared understanding is a procedural condition of collaboration.
- A shared perceptual field (through the settings) enables a joint attention to an object of reference; thereby, it facilitates the establishment of shared understanding in collaboration.

5.4.1 Scenario

The scenario simulated typical communicative situations where people rely on pointing such as introducing people, giving guidance and negotiating. It drew on theoretical views of deixis and observational insights of gestural communication.

5.4.2 Design Instruction

A test set-up was designed and implemented (Figure 15). The researcher formulated a design instruction[94] to guide the implementation of a test bed, an experimental augmented videoconferencing system: it had to provide an overlap of the participants' perceptual

[94] The first test bed was implemented by Guido Kuehn, the latter one by Swen Walkowski from Wiesbaden University of Applied Sciences.

fields in order for them to have common ground for inferring shared understanding of a deictic reference[95] (Figure 9).

5.4.3 Conceptual Design and Implementation of a Test-Bed

The test bed[96] was an augmented videoconferencing system that had capability

- to track the direction of pointing relative to the screen indicated, and
- to visualize the point of reference in live video stream.

The application engineer came up with the idea to provide two displays at each end: one for the local site view, and the other for the remote site view (Figure 15). Additional features were required to capture the direction of a pointing gesture and for visualising the indicated point in the video stream; they included a pointing device (WiiRemote), infrared detectors (the deictic origo -> the reference object in a video frame), and Augmented reality (AR)-imagery (Figure 17).

In the implementation phase, the researcher was playing a user role in a dialogue with the application engineer.

In the first test bed, there were four different modes mapped on the pointing device (Figure 16). The first, 'plain' pointing was for indicating objects. For the human objects, there was a spotlight-like representation, and an arrow for the nonhuman objects (Figure 17). It was also possible to expand/shrink a circle and to turn the arrow anti/clockwise, even to make it spinning.

Information mode was a pointing mode to bring visible additional information (Figure 16, right) when a label (fiducial marker) of a remote participant was pointed to. It was meant to support the chairperson of a videoconference should s/he forget the name of the participant. The third mode enabled annotations, designed to support group decision-

[95] In the beginning, the application engineer of the first test bed and the researcher did not have lingua franca to communicate properly: the former had background in Computer Science, the latter in Architecture. As 'perceptual field' did not make much sense to the application engineer, the researcher reformulated the instruction in set-theoretical terms by drawing two intersecting circles on a piece of paper and describing the origo of a circle as the spatial perspective of an observer, and the circle around the origo as a perceptual field opening from that perspective. In order for the two participants to have shared perceptual field, the two circles have to intersect. In order for both of them to perceive the same object, it has to be within the intersection of their perceptual fields. Since then, the idea was clear. The application engineer used the word 'information space' when referring to perceptual field. Regarding the question '*how*', the researcher told him to feel free to use 'whatever way works'.

[96] For the technical specification of the first test bed, see Kuehn & al. (2007).

making. Also the fourth option, a snapshot[97], was meant to support negotiations. The chairperson could take pictures of various suggestions and flash them (max. 4) side-by-side before the group made a choice between them.

It appeared in the pilot test that a few minutes' introduction was too short for the test users to become familiar with many pointing modes. Therefore, the second test bed, designed for *two-way pointing*, featured only a plain pointing mode. As there was then a pointing device available at both ends, simultaneous pointing and related co-operative/competitive phenomena could be investigated (Figure 30).

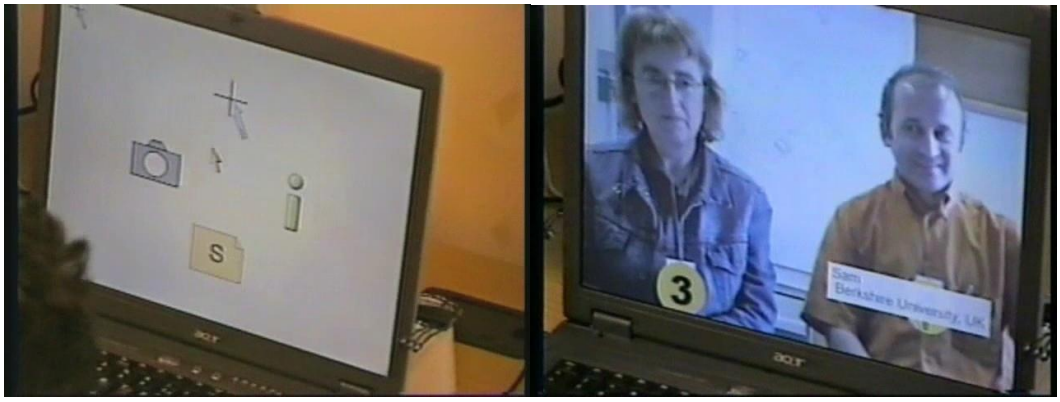


Figure 16 The first test bed provided four pointing modes
When the information mode was used in pointing to a label (fiducial marker) worn by a participant, the role name and affiliation appeared in the video stream (the right frame).

Figure 17 Pointing representations for human (spotlight) and non-human objects (arrow)[98].

[97] Sound effect was included because audio feedback is required in order for the whole group to register at which point picture is taken.

[98] The video frames to illustrate method/findings are from the tests run on both versions of the test bed, the first with G. Kuehn and the latter with S. Walkowski.

Table 6 Methods applied in Case study 1

<table>
<tr><td>CASE STUDY 1:
EXPERIMENTAL DESIGN</td><td colspan="3">Hypothetical assumption:
By design of the settings (>interaction space), it is possible to restore the communicative function of an across-site pointing gesture in the video-mediated communication.</td></tr>
<tr><td>DESIGN INSTRUCTION
to implement
a test bed</td><td colspan="3">(Drawing on multidisciplinary research literature & ethnographic observation)

See Figure 6</td></tr>
<tr><td>TEST DESIGN</td><td>TEST SETTINGS
(In collaboration with software engineer)</td><td>PROCEDURE
TEST TASKS</td><td>TEST BED
(implemented by software engineer):
-2 screens at both ends
-Pointing device
-IR-detectors (to capture the point indicated)
-AR to visualize the point indicated</td></tr>
<tr><td rowspan="2">USABILITY TEST</td><td colspan="3">Test sessions for one-way pointing on test bed 1, implemented by G. Kuehn
Test sessions for two-way pointing on test bed 2, implemented by S. Walkowski
Test groups consisting each of 1 to 3 local and 2 to 4 remote participants</td></tr>
<tr><td colspan="3">FIVE COLLABORATIVE TEST TASKS:
Each task required across-site pointing gestures in order for the group to complete a joint task in a successful way.
1. Introducing people.
2. Taking questions from both sites.
3. Suggesting changes to a drawing.
4. Pointing to an item located at the remote site.
5. Guiding the way on a map located at the remote site.</td></tr>
<tr><td>VIDEOANALYSIS
(=Principal source of evidence)</td><td colspan="3">Video data from the usability tests:
1. To identify situations where participants use pointing gesture.
2. To judge from the addressee's response (verbal and/or gestural confirmation) whether or not an across-site pointing is guiding the addressees' attention to the object of reference
3. To identify whether the participants successfully complete the joint tasks.</td></tr>
<tr><td>EVALUATION</td><td>-DO THE SETTINGS ENABLE GUIDANCE OF ATTENTION IN ACROSS-SITE COMMUNICATION?</td><td>-DO THE TEST USERS ESTABLISH SHARED UNDERSTANDING OF THE OBJECT OF POINTING GESTURE IN ACROSS-SITE COMMUNICATION?
-Do they adapt to the context of interpretation?
-Do they rely on the support provided by the experimental settings for gestural communication?</td><td>-VIABILITY OF THE CONCEPT
-Does the design enable the communicative function of an across-site deictic gesture?</td></tr>
<tr><td>NARRATIVE DATA
from
Feedback discussion & Questionnaire
(Anecdotal evidence)</td><td>-User perception of the settings:
-Acceptability in terms of ease/difficulty to orient / navigate in the experimental setting</td><td>-User perception of experience:
-Perceived
- (un)helpfulness
- (un)convenience</td><td>-User perception of the design:
-advantages
-disadvantages
-shortages
-suggestions for technical improvement</td></tr>
<tr><td colspan="4">Summarizing the findings from Case study 1 → Moving on to CASE STUDY 2</td></tr>
</table>

A two-colour outline made the arrow visible against any colour background. The representations were also colour-coded according to the site. Furthermore, it was possible to use 'mediated' pointing both for remote and local objects (Figure 27)[99]. The representation for a human object was a bluish spotlight, and an arrow for non-human ones.

5.4.4 Test Settings

The test settings (Figure 15), the procedure, and five test tasks were designed (Appendix 4). The test setup was modified from Clark's & Wilkes Gibbs's (1986) design for a study of a collaborative process (=referring as a collaborative process). It consisted of two PCs, a pointing device and two displays for each site to provide both a local and a remote site view for the participants.

The participants of a test were located in the same room though separated by a partition (Figure 15). They shared audio co-presence but could not see to one another's side.

In the first test bed, the opposite site computers were connected by a cable. The context of action and interpretation was thereby simplified, to avoid problems by delay and video quality.

There was only one pointing device available in the first set of tests, whereas in the tests for two-way pointing, the participants had one pointing device at both ends.

The test tasks were printed on an A4 for each participant; another page described the functionalities mapped on the pointing device (Wii Remote).

5.4.5 Test Tasks

The test tasks (Appendix 4) simulated natural communication where people frequently rely on gestural deixis. The tasks were to yield relevant data to determine whether the communicative function of an across-site pointing gesture is successfully mediated through the settings. Relative freedom was left for the test users' spontaneous moves.

- The first task simulated a gathering where someone had to introduce people to the other participants.

[99] Face detection (in Test bed 2) did not function up to expectation as it misinterpreted several times the object. The problem is not discussed in more detail in the thesis because the focus is not on the *type* of pointing representation but whether or not pointing directs the attention of an addressee.

- The second task simulated a meeting or a classroom. One of the participants had to play the role of a chairperson who was controlling the floor, or, a teacher taking questions from the students.
- In the third task, people had to give operational instructions to a person who was at the opposite site. Demonstratives such as '*that* one' frequently appear in co-occurrence with pointing instructions.
- The fourth task focused on negotiation. People usually point for instance when they have to explain something on a drawing: "Can we move the wall further there to get space for one more group here?"
- In the fifth task, someone had to be guided to a destination that was unfamiliar to him/her. People typically indicate with a hand where to go, or point with a finger the directions on a map.

5.4.6 Usability Evaluation

Test sessions due to last maximum 30 minutes were run on two versions of a test bed. The test users were young university students or members of staff, familiar with Computer Mediated Communication; many of them had prior experience of video-conferencing. The age range was from the early twenties to the sixties.

Table 7 Test sessions and test users /session

Test sessions	**Test bed 1 one-way pointing**	**Test bed 2 two-way pointing**
1 (Piloting)	2+2	3+3
2	2+2*	3+3
3	2+2*	3+2
4	2+2*	3+3
Additional tests	1+4/3*	
In total	20	23

*software engineer included

It turned out in the pilot session that several different pointing options would be too much for a novice user to cope with; therefore, the test design for the test bed 1 was modified so that the software engineer was as one of the participants (who mainly used pointing functionality) and consequently, the focus of analysis was in the first set of tests on the addressees' responses. Most sessions for one-way pointing were run with two participants at both ends. The sessions started with ten minutes' introduction where the purpose and

the procedure of the tests were explained, and where the software engineer explained the system and instructed how to operate it. The test users were given each a role name and affiliation (label). Each participant also signed the consent request[100].

The test sessions for two-way pointing were run with three participants at both ends[101]. The participants had a few minutes' time to use the pointing device before they took their positions and started the tasks (Appendix 4).

The test sessions were video-recorded with two cameras. After each session, a brief feedback discussion followed. The participants also filled in a questionnaire form (Appendix 5, Appendix 6).

5.4.7 Analysis of Data

The data collected from the tests through observation, video recording, feedback discussions and questionnaire focused on spatial navigation and communication strategies (=social-cognitive navigation).

Principal evidence was sought from the video recordings (=hard data in the pictorial form) to determine *whether or not the settings enabled effective gestural deixis in the across-site communication* (whether or not the participants managed to infer *shared understanding* of a referent when a person indicated an object located at the remote site (Chapter 6.1). Video recordings also provided pictorial evidence of whether or not the test users adopted pointing functionality in their communication (Figure 18, Figure 27; Chapter 6.2).

Furthermore, they showed whether/how the users adapted to the test settings (>an unfamiliar frame of interpretation, Chapter 6.3). To determine the communicative (in)effectiveness of a pointing gesture (Figure 18), attention was paid to behavioural signs such as the addressee's gaze following, nodding, verbal confirmation, requests for additional information, and to the outcome of a joint task in progress.

After both sets of tests, analysis of the data was carried out in a daylong session, in the presence of the researcher and the application engineer. The researcher explained the method, and the session started by identifying situations where across-site pointing

[100] Options for using video frames as illustrations of methods / findings: 1. full consent, 2. consent given if face smudged; 3. consent declined.

[101] According to Nielsen (1995), five test users is enough /economic in the type of tests where the point is to gain basic insights of usability problems. Such tests help to identify more specific questions to focus on in further development and testing. The number is too low for proper statistical evaluation.

occurred. Such *key episodes* (e.g. Figure 18) were further investigated by replaying videos - some passages over and over again – while the researcher focused on the following questions:

- *for what* (spatial-navigational/communicative) *purpose* the participant then used pointing,
- *how* s/he pointed (finger pointing vs. pointing with WiiRemote),
- what impact pointing had on the co-participants' communicative acts, in particular,
    - whether the addressee's gaze direction (=attentional focus) turned to the indicated object,
    - whether the addressee(s) nodded / uttered a sound for confirmation,
    - whether the addressee's verbal response indicated shared understanding, and
    - whether the addressee managed to carry out a task according to his /her (=a remote-site speaker's) guidance.

Figure 18 Ineffective (left) **and effective** (right) **pointing gesture in the across-site communication**

The feedback discussions and the questionnaire provided narrative data from the user perspective. It was organised into three main categories (Table 6), concerning

- ease / difficulty to orient in the experimental settings,
- helpfulness / unhelpfulness and convenience / inconvenience in communication,
- advantages, disadvantages, technical problems and ideas for the improvement of the application.

It was then possible to align the test users' views with the findings from video data, and thereby, to anchor the conclusions on a firmer basis to observation, video analysis and to the informants' self-reported data.

5.5 Case Study 2: Hybrid Interaction Design for Video-Mediated Lecturing

Case study 2 (Table 8) was *a practical design case* where a lecture theatre (Figure 19) was spatially orchestrated to support video-mediated interaction and communication across geographically distributed locations.

The point of departure was in the venue (>the spatial and technical constraints of the lecture theatre), and in the users' experience / action (>the user perspective).

The notion of deixis guided the coordination of the opposite site role perspectives and thereby, the facilitation of a joint attention to an object of reference, by means of

- reciprocity of perspectives (> mutuality as a functional connection), and
- referential triangle (> deixis as an act of reference) (Figure 9, Figure 10).

Case study 2 started with a site visit to the lecture theatre to be turned for video-lecturing and to the other universities involved in the project. Initial communication with stakeholders took place. At the local site, a role play was arranged with the Head of AV to map the main constraints of the lecture theatre for design and also, to consider requirements both from the teacher's / student's point of view. The method to map requirements for the settings from the teachers' point of view, focus group discussions were arranged at the local and remote sites. Furthermore, lectures were observed at each participant university to better understand the teacher-student interaction during lectures in the subject matter. The specification of their role requirements were taken into account in the implementation. After AV staff's technical testing of the system, user feedback was collected in a showcase video-lecturing session, and video-lecturing rehearsals were arranged for teachers.

5.5.1 Object of Design

The aim of Case study 2 was to design functional affordances for gestural deixis in the hybrid interaction (case: video-mediated lecturing).

Table 8 Methods applied in Case study 2

CASE STUDY 2	**TEACHER-STUDENT INTERFACE FOR VIDEO-LECTURING** **Designing interaction space for video-mediated lecturing**
SITE VISIT	**To map constraints of the site due to be converted for video-lecturing:** • **Snapshots** • **Communication with stakeholders (AV, teaching staff)**
ROLEPLAY	**In the lecture theatre due to be converted:** • **AV professional in the role/position of the teacher** • **Researcher in the role/position of a local student** **Reflecting upon a set of different presentation/interaction scenarios and their implications to spatial and technical arrangements.** • What needs to be seen / heard / done from the teacher's perspective? • What needs to be seen / heard / done from the student's perspective? • What needs to be available / at hand?
FOCUS GROUP DISCUSSION	**Initial mapping of the role requirements for the settings**
ETHNOGRAPHIC OBSERVATION	**Mapping teaching practices in the subject field:** • Attending a set of lectures at the participant universities • Observing regularities (>patters) and differences (>range) in the use of space / use of tools / way of presentation to / interaction • Notes taken, and (if consent given) snapshots + a few seconds videos shot • Observation report checked with the observed teacher to confirm the interpretations
SPECIFICATION OF THE ROLE REQUIREMENTS >>> DESIGN INSTRUCTION	**DESIGN INSTRUCTION FORMULATED FOR TEACHER–STUDENT INTERFACE** • **Multi-site audience from the teacher's perspective** • **Teacher-student interface from the local/remote student's perspective**
(IMPLEMENTATION)	Providing reciprocity of perspectives / referential triangle in the implementation
EVALUATION **Feedback from test sessions and showcase**	**Observation of test sessions / showcase sessions / video-lecturing rehearsals for teachers** **Feedback questionnaire after the showcase:** **User ratings of** • **functional aspects of video-mediated lecturing (Likert scale)** • **communicative aspects of video-mediated lecturing (Likert scale)** **User comments from the showcase and teacher rehearsals**
EMERGING PRACTICE	**Observation of video-lecturing in progress** • **Learning from emerging practice for further design**
METHODOLOGICAL OUTCOME	**REFINING A DEIXIS-BASED DESIGN APPROACH →** **LOCUS-MODUS-TOPIC-FOCUS FRAMEWORK FOR WORKPLACE DESIGN**

In workplace design, it is not necessary to explain every aspect in the emergence of the communicating participants' shared understanding; what is then relevant is the question

how the communicative perspectives are coordinated in order to facilitate functional affordances for (topic relevant) deictic practices.

Communicative situation is the spatial interface that connects the participants' perspectives. The participants' interaction space is their situational common ground and information conveyor: people employ resources that are situationally available to them. The situation is, in that sense, a time-slice and affordance[102] for perception and action (>collaboration).

When a communicative situation is taken as a point of departure, we focus on *mutuality* and *reciprocity of perspectives*. To put in design-relevant terms, communication is the participants' multimodal, jointly constrained and coordinated navigation. Multiple coordination nodes underlie reciprocity of perspectives: First, *the human body* is the node of sensory-modal coordination. Second, *the position* of a person (place[103]) determines what information is available for the person's *perception and action* at a particular time; at the same, it provides the spatial origo of the person's intentional action. The third design-relevant node is a joint one, *the topic* of the social encounter[104]: it moulds the participants' contributions within the settings[105] into a (more or less) coherent whole, that is, what is jointly getting shape, whether a conversation or something else. The fourth node, individual cognition, provides the participant's motivational frame for the situational focus of his/her communicative acts (Figure 10, Figure 11).

From the modal point of view, the coordination of participant perspectives is particularly complicated when hybrid settings are designed: the spatial layout of the (local) premises *and* the technical connection (audio, video) to the remote site have to be adequately arranged in order to facilitate effective across-site communication and collaboration.

The findings from Case study 1 had (by then) confirmed the initial assumption that the notion of deixis is an indispensable tool for spatial design if situated in a communicative situation: it then links the spatial (>position), modal (>audio, video), social (>participants' perspectives) and cognitive (>perception/ intentional action) aspects of communication; thereby, it resolves the complicated problem in the spatial orchestration

[102] Concept originally coined by Gibson; later adopted within HCI-community, see e.g. Hartson, 2003.
[103] *Place* stretches across object and its environment (e.g. Aristotle, 2007; Tuan, 1974, 1977; Norberg-Schulz, 1980; Harrison & Dourish, 1996; Fitzpatrick et al, 1996; Ciolfi, 2004; Dourish, 2006). *The Vitruvian man* by Leonardo is a classical depiction of a human-environment interface.
[104] In a design process, *topic* refers to *interpretation of practice* as it is understood / formulated in the project.
[105] In this regard, ICT-mediation is just a modal feature because it does not make any foundational difference to the human-human interface.

of the complementary role perspectives for hybrid interaction and communication. In the design of the settings, the topic activity has to be considered in *functional* terms.

Consequently, the following rationale guided the conversion of the lecture theatre (=>hybrid settings):

- *Functional affordances for shared understanding result from the coordination of the participant role perspectives, through the settings, in a topic relevant and task sensitive way.*

5.5.2 Location as Point of Departure in Design

The design case was a traditional lecture theatre, with ascending rows of seats for ca 150 persons, the entrance being on the left of the audience (one in the front, one in the back). The theatre was furnished with black boards and a white board in the front (Figure 19).

The main spatial and technical constrains for possible changes were mapped with AV staff, keeping in mind the following questions:

- How to facilitate natural interaction between the local and the remote site?
- How to provide relevant views and adequate feedback for / from different role perspectives?
- How to extrapolate the guiding / disambiguating / supportive cues (such as hand gesture) and floor control functions to video-mediated situation?

More details were mapped in a role-play with a senior AV-professional. The role of the teacher was allocated to the AV-professional who positioned himself in the front, as if giving a lecture. The researcher took a position in the auditorium prompting scenarios and playing the role of a student. Kind of a cognitive walk-through of a lecture followed, along the lines:

- *Let's assume that you are the teacher and you have the local and remote audiences there, so what do you need 1) to see, 2) to hear and 3) to do in order to present your lecture and interact with the students?*

Apart from mapping spatial and technical constraints, ideas were created and discussed, focusing on the spatial layout and optimal positions of the cameras, displays and microphones. The aim was to provide an effective and user-friendly learning interface.

Furthermore, the settings ought to be also easy to assist and maintain by the technical staff. In order to specify complementary role requirements for the settings, data had to be collected from different user perspectives.

Figure 19 Lecture theatre as a point of departure for the rearrangement

5.5.3 Action / Experience as Point of Departure in Design

The teacher's and the student's perspectives are the two role perspectives relevant to lecturing. Another relevant stakeholder group is AV staff; they provide technical assistance and maintenance, and they give helping hand in problem shooting. In the following, the design process is described focusing on the teacher's user perspective (=role requirements for the spatial settings). The topic activity was lecturing/learning in Bio-Medical Sciences.

A meeting was organised at each participant university with the members of the teaching staff due to start video lecturing. Their role requirements for the design were initially mapped in a focus group discussion (Appendix 7) and teacher observations.

Ethnographic observations of lectures in the subject field were carried out at each participating university to gain understanding of different settings, institutional cultures, and individual lecturing styles (Figure 20, Figure 31).

A number of teachers (10) were observed to identify spatial choreographies and communication strategies in the interaction with the students (Figure 33):

- how they use space in the classroom
- what kind of resources / strategies they employ,
- how they communicate the contents, and
- how they interact with the students.

Snapshots were taken and short videos were shot to capture accurate data for analysis[106]. The observation notes were summarized in a report that described the structure of a lecture, the teacher's spatial/communicative performance, and the resources used for communicating the contents to the audience. Particular attention was paid to presentation and interaction (Figure 32, Figure 33).

Furthermore, assumptions were made of the teacher's pedagogic strategy, why something was done in particular way. The observation report was validated in a discussion with the informant. The researcher was then able to check whether her initial interpretations were on the right track. The session was also for mutual learning because the informant had an opportunity to analyse and to learn from his/her lecturing in the light of pictorial evidence.

5.5.4 Reciprocity of Perspectives as Challenge for Design

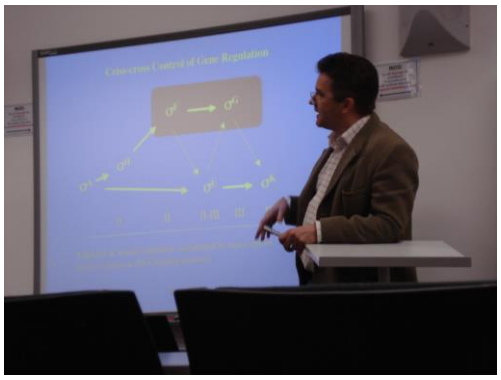
A multi-site collaboration poses particular challenge for the settings in terms of gestural deixis. From the teacher's point of view, a multi-site audience is complicated because s/he has to share his/her equal attention to local and remote participants. Furthermore, s/he needs visual feedback.

The observations show that there are *three* principal *reference points* in the teacher's presentation: they look at their notes, cast a quick glance at the audience, thereafter turn to the slide (in front of the class), indicate something on it, and then turn to look at the audience again (Figure 20, Figure 31).

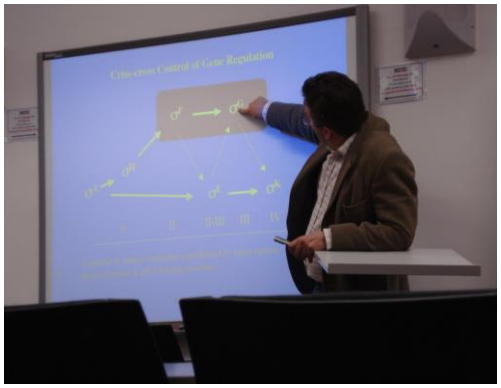
[106] In case taking pictures/shooting videos during observation was agreed upon.

Lecture observation

Gaze contact with the audience:
to monitor the overall situation,
to maintain the mutual contact.



Turning to the smart board for the
focus of topic on the screen:
guiding attention of the audience
to the object of reference.



Indicating a point with the finger:
disambiguating the referent.

Gaze contact with the audience:
seeking feedback to assess whether
the audience has got the point:
assumption of shared understanding.

Figure 20 Teacher's perspective
Lecturing was observed as an activity unfolding between the teacher and the students. Attention was paid to typical patterns of communication and interaction.

The lecturer has to design his/her talk to be compatible with the audience's reception, and such nonverbal procedure helps him/her by informing whether the communication 'sinks in', or, whether s/he has to carry on explaining. One of the requirements is that the teacher has to see the *whole* audience with a quick glance (Figure 34).

The Question and Answer sessions (Q&A) are part of the teacher's pedagogic strategy. Interaction may take place at the end of the lecture, but some teachers prefer to pause after every package they explain to make sure that it has been understood (Figure 33).

Both the local and the remote students should see equally well what the teacher is explaining, and how s/he is doing it; they should be able to capture the lecturer's gestural (>procedural) cues, and to catch his/her attention should they have any problem or a question in their minds. During Q&A sessions, they may also wish to see who is in voice at the remote site.

The problem is how to coordinate complementary role perspectives of collaboration across different sites in order to facilitate

- joint attention to an object,
- across-site social presence, and
- interaction across different sites?

The findings from Case study 1 (Chapter 6) suggested that *the modal scope* of an interaction space, with regard to the participants' common goals of collaboration, constrains what can be done in such settings (effectiveness of collaboration). The modal scope not only refers here to the geometric aspects and to the boundaries of the interaction space but also to the availability of information (Kuusisto, 2004)

- from the surrounding environment,
- situated in the participants' perspectives (e.g. Roth, 2001; Gabora, 2008), that is:
    - background knowledge (worldview) of the participant in relation to
        - the on-going communication (the topic), and
        - one another's background knowledge and motivation.
    - practical ability to apply background knowledge to ongoing situation.

The basic requirement is a functional connection (>audio, video) between the local site and the remote site. In addition, the technical quality of video/audio has to be high enough in order for both the local and the remote audience(s) to see/hear properly the object explained.

Each modality (audio, video) has to be technically considered and spatially orchestrated in order to provide (a topic relevant) reciprocity of (the role) perspectives and thereby, to enable an across-site communication that is both effective and cognitively effortless.

Apart from the controls and their positions, *who* operates the control (for instance, a relevant camera view[107]) has to be considered taking into account that the teachers tend to move quite a lot while lecturing.

In addition, a mode switch from delivering the contents to taking questions has to be both technically and cognitively effortless (Figure 32, Figure 33): there is no time for wasting because the lecture is always heavily packed with information.

Though Case study 2 did not include the spatial arrangements at the remote site, they had to be taken in the implementation into account in order for the settings to facilitate effective across-site communication.

5.5.5 Collecting Feedback from Users

Before the video-mediated lecturing 'went live' in the renovated lecture settings, there were a few technical and showcase sessions[108]. A questionnaire was used to assess collect feedback from the showcase participants; they were asked to assess the showcase performance in comparison to traditional collocated lecturing rating technical quality and fluency of performance and communication according to Likert scale (Table 9, examples in Table 10). Furthermore, additional user comments were collected which revealed both positive impressions and concerns:

- *…the fact that the equipment from a technical point of view worked.*
- *I found the lectures were nearly the same as lectures in traditional co-located settings.*

- *I left concerned that this system is not sufficiently robust to enable the delivering series of lectures without breaking down at times.*
- *A bit stilted due to lack of practice.*

The findings from the showcase then informed teacher rehearsals. The members of staff who were due to start video-lecturing could there familiarize with their novel frame of action / interpretation, both from the teacher's and from the student's perspective: they gave a short lecture and attended others given by their colleagues. Particular attention was

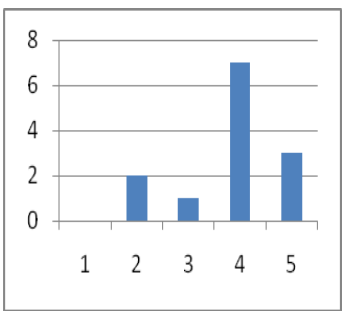
[107] Alternatively camera tracking
[108] Arranged by the Swan Project Coordinator

paid first, to adequate ways of guiding the attention to the audience in a context-adequate way, and secondly, to generating a sense of social presence at the remote site (Figure 35).

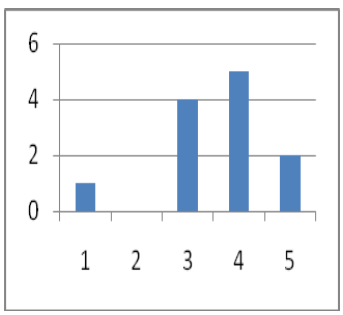
Table 9 Feedback questionnaire (refined)

	RATING YOUR EXPERIENCE	FROM	1	2	3	4	5	TO
1	Quality of sound transmission: easy to hear	Below satisfactory						Very satisfactory
2	Quality of video transmission: easy to see	Below satisfactory						Very satisfactory
3	Quality of slides transmission: easy to see and read	Below satisfactory						Very satisfactory
4	Cursor movement	Below satisfactory						Very satisfactory
5	Blurring, pixilation of slides	Highly disturbing						Not at all disturbing
6	Background noise, echoing	Highly disturbing						Not at all disturbing
7	Technical interruptions	Highly disturbing						Not at all disturbing
8	A-synchronicity	Highly disturbing						Not disturbing at all
9	Camera view = focus and scope guide attention properly	Below satisfactory						Very satisfactory
10	Level of lighting to see lecturer's face properly	Below satisfactory						Very satisfactory
11	Sense of presence of the lecturer talking to you	Below satisfactory						Very satisfactory
12	Sense of presence of the remote audiences	Below satisfactory						Very satisfactory
13	Interactivity of Q&A sessions	Below satisfactory						Very satisfactory
14	Fluency of Q&A sessions	Below satisfactory						Very satisfactory
15	People involved in Q&A conversat from the other sites	Difficult to follow						Easy to follow
16	Floor control (turn taking during Q&A)	Below satisfactory						Very satisfactory
17	Your experience compared with attending a traditional lecture	No difference						Very different
18	The most positive aspect in the videoconferenced lecture?							

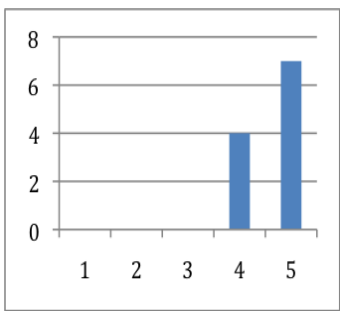
Table 10 Feedback from the showcase, extracts



Video quality



Camera view (>attention)



Presentation C

Technical quality was assessed along the scale below satisfactory (1) - very satisfactory (5). Communication was assessed along the scale difficult to follow (1) - easy to follow (5).

5.5.6 Emerging Practice

To conclude the project, a few more observations were conducted of emerging practice (Figure 35). Furthermore, guidelines were drafted for accumulating best practices and tips for problem shooting for the benefit of novice users.

5.6 Summary

The framework for workplace design (Chapter 4) was evaluated in two case studies. The chapter explained how situated deixis encompasses, as a tool for spatial design, the two key conditions for effective synchronous communication and collaboration, that is, *reciprocity of perspectives* and *referential triangle*.

Case study 1 explored gestural communication over a video link. Using deixis as a conceptual tool in the spatial design, an experimental test set up was implemented where video-mediated pointing could be investigated. The assumption was that such settings facilitate the communicative function of an across-site pointing gesture. Usability tests were run with four to six test users on two versions of a test bed. Data collection included observation, video recording, post-test discussion and questionnaire. Due to the purpose and type of the test (viability of concept), quantitative methods had only a minor role. Principal evidence was sought from video data to determine whether the settings enabled/supported the participants' joint attention and shared understanding of (gestural) deictic reference. Thereby, it was established whether or not the design instruction was viable and the assumption underlying it on the right track. Observations and narrative data were used to extend / strengthen the basis for interpretations and conclusions.

In Case study 2, *a method for the coordination of the opposite site role perspectives* in the video-mediated collaboration was described. In a conversion of a traditional lecture theatre, situated deixis guided the orchestration of the role perspectives *to facilitate functional affordances* for video-mediated lecturing to a multi-site audience. Requirements for the settings were specified through focus group discussion and observations. Preliminary user feedback was collected in a showcase before the settings were introduced into regular use.

6 Findings

This chapter summarises the main findings from the two case studies that were conducted to validate the proposed framework (Chapter 4) for spatial design. Findings from the usability tests (Case study 1: experimental settings for gestural deixis, Table 11) are presented in Chapter 6.1-6.4, and findings from the coordination of complementary role perspectives for video-lecturing (Case study 2: hybrid interaction space for teacher-student interaction) in Chapter 6.5-6.7.

In Case study 1, it was first examined from video data whether or not the test settings provided functional affordances for effective (pointing) communication, in other words, whether they supported the test users' guidance of attention, joint attention and formation of referential triangle (speaker - addressee - object of reference) thereby, shared understanding of a deictic reference. To assess usability of the settings, the users' adoption of the application as well as their ability to orient in its artificial spatial frame were analysed. The evidence was based on video data and confirmed by narrative data (user comments).

In Case study 2, it was established drawing on observation data and narrative data (user perceptions) from preliminary use whether the lecture theatre enabled reciprocity of perspectives over a video link, and thereby, provided functional affordances for video-mediated lecturing.

In a theoretical consideration, the findings from the two case studies concern design principles for hybrid interaction space. They are discussed in the summary (Chapter 6.8).

6.1 Effectiveness

Effective communication is defined in this context as follows:

An across-site pointing gesture (across-site gestural deixis) is in the communicative consideration effective if the participants of communication are able to locate/identify

- without additional verbal explanations, and
- independent of whether or not the object *and* the person pointing to / referring to it is local or remote,

who is pointing to and what is pointed to, and thereby, to establish shared understanding of the object of gestural reference. In such a case, the participants of communication are able to coordinate their perspectives and contributions in a successful manner, and their

multimodal communication is, in the communicative-functional consideration, effective and coherent.

Assumption: The assumption underlying the design of interaction space was that in order for the participants of communication to be able to establish shared understanding of an object of gestural reference, they have to have a visual perceptual common ground (p.84, Figure 6).

The experimental settings were designed to enable the communicative function of a pointing gesture in the across-site communication; in other words, the participants were due to see - each from their own perspective - who is pointing and what is pointed to. In such settings, the participants were assumed to coordinate their mutual perspectives and to cognitively re-integrate *the spatially fragmented referential triangle* between the person pointing, the object pointed to, and the observer (Figure 9). As a result, it was assumed that the test users would establish *a joint attention* to the object of gestural reference. Joint attention is a precondition for *shared understanding* which again is a step in the *coordination of contributions* (Table 6).

Method and outcome of analysis: The question whether the participants established joint attention / shared understanding and successfully carried out joint tasks was investigated with the help of test tasks that required across-site pointing (Table 11). In the analysis, communicative situations where across-site pointing occurred were identified from the video recordings of the sessions. Attention was paid to the communicative impact of pointing:

- Does the addressee's gaze start to follow the point indicated by the speaker (Figure 22)?
- Do the participants' gestural and verbal responses indicate shared understanding?
- Do the participants successfully carry out (across-site) co-operative tasks?

Coordination of perspectives → joint attention: It was possible to analyse from video recordings of the test sessions where the participants were looking at any point of the task (Figure 22). If someone at the local site was pointing to objects at the remote site, the addressees were keen following the line of pointing in the video stream. However, looking at the same point does not yet guarantee that the perspectives were coordinated in a communicative regard: people may just coincidently have focused visual attention on the same point, without any common referent or interest. Only if their gazes follow the line of pointing in the same manner, it becomes clear that they have a common object in

the centre of their visual attention. For a referential triangle, even more evidence is needed. The participants have to understand, each from their particular spatial and communicative perspectives, *who* is pointing to *what*: in other words, they have to be aware of one another's presence, too.

Referential triangle → shared understanding: The question whether or not the participants were able to cognitively re-integrate the referential triangle, was examined in a cross-modal way. Attention was paid to the contents of *verbal* communication in concert with *visual* information. To prove that the referential triangle has been successfully re-integrated, the following kind of situations had to be found in the data:

- A test participant uses personal pronoun *you*, while pointing with the pointing device to the video image of a remote person without mentioning his/her name and the person pointed to among the remote participants responds.

Drawing on the video data (e.g. Figure 21), it was possible to show that people at the remote site were, indeed, able to locate/identify the referent. Similarly, if a participant pointed to an object at the remote side wishing to know what it was (e.g. Figure 25), people at the remote site had no problem in providing a correct answer:

- *How are you?* (pointing to a remote participant)
- *I am fine, I am fine!*

- *I was wondering… what is that?* (pointing to a remote object)
- *Oh, that - it is a paper box.*

Figure 21 Recognition of the referent
The system tracks the direction of pointing and visualises the point indicated as a spotlight in the video stream. The right person responds by taking the floor. The person next to the addressee also focuses attention to him/her.

Table 11 Main results from the usability tests

	A Joint attention	B Shared understanding	C Spatial settings
ASSUMPTION	The experimental settings provide common ground in the visual modality for both end participants: they share identical views of both ends	The test users are able to link who is pointing to what is pointed, due to their visual 'common ground' and AR-visualization of the point indicated. => they identify the object of gestural reference	Due to visual common ground and AR-visualisation, the experimental settings make possible for the human cognition to (re)integrate a referential triangle to establish a joint attention, and thereby, to establish shared understanding of the object of gestural reference without verbal explanation.
METHOD / VIDEO DATA	**USER TESTS with 5 test tasks (See Table 5); two versions of a test bed: (one-way, two-way)** 1. To identify situations where the test participant points to an object that is located at the remote site 2. To register whether or not the addressee's visual attention turns to the object indicated. 3. To register the addressee's related verbal / gestural / paralinguistic responses 4. To register the addressee's consequent moves / actions 5. To conclude whether or not the participants manage to carry out co-operative actions and carry out a joint task.		
QUESTION	**DO THE TEST USERS MANAGE TO GUIDE THE ADDRESSEE'S ATTENTION TO THE OBJECT OF (GESTURAL) REFERENCE?**	**DO THE TEST USERS ESTABLISH SHARED UNDERSTANDING OF OBJECT OF REFERENCE?**	**IS THE CONCEPT / DESIGN VIABLE?**
FINDINGS	**YES**	**YES**	**YES**
	The participants turn to look at the point indicated **-COORDINATING PERSPECTIVES →** **-JOINT ATTENTION**	**The participants identify the objects pointed to by a remote participant.** **- REFERENTIAL TRIANGLE →** **- SHARED UNDERSTANDING**	**The participants successfully carry out co-operative tasks.** **COMMUNICATIVE FUNCTION OF POINTING GESTURE ENABLED**
CONCLUSION OF VIDEO ANALYSIS	THE SETTINGS PROVIDE SHARED INFORMATION SPACE IN THE VISUAL MODALITY, (=A PRECONDITION FOR JOINT ATTENTION) THEREBY, THEY PROVIDE VISUAL COMMON GROUND FOR GESTURAL DEIXIS OVER A VIDEO LINK	DUE TO AR-VISUALISATION, THE SETTINGS SUPPORT THE FORMATION OF A REFERENTIAL TRIANGLE >>> THE PARTICIPANTS ARE ABLE TO ESTABLISH SHARED UNDERSTANDING OF THE OBJECT OF REFERENCE THEREBY, THEY SUCCESSFULLY COORDINATE THEIR CONTRIBUTIONS.	THE HYPOTHETICAL ASSUMPTION GUIDING THE DESIGN (>REFERENTIAL TRIANGLE IN THE VISUAL MODALITY) IS 'ON THE RIGHT TRACK'. THE COGNITIVE PRINCIPLES UNDERLYING SHARED UNDERSTANDING SUCCESSFULLY GUIDED THE SPATIAL DESIGN AND IMPLEMENTATION OF THE SETTINGS
COMPARISON	**FINDINGS FROM VIDEO DATA CORRELATED WITH NARRATIVE DATA**		
USER PERCEPTIONS NARRATIVE DATA	Ease/ difficulty - to orient /navigate in the test settings.	Perceived -un/unhelpfulness - in/convenience	-Advantages /disadvantages -shortages / suggestions how to improve the design.

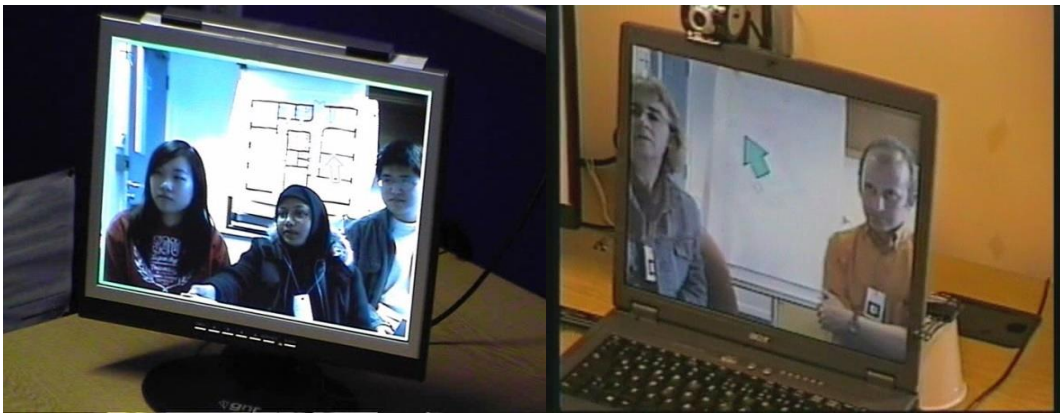


Figure 22 Gaze direction indicates the communicative impact of pointing
The gaze follows the path of pointing.

Use of demonstratives and locative adverbs → multimodal deixis: As demonstrative pronouns and locative adverbs link the semantic and practical dimensions in the speech situation, they helped to find out whether or not the participants of communication shared a visual common ground. To give an example, an instruction to move a table *there* completely depends on the speech situation; otherwise it does not make practical sense.

When the participants were negotiating with the help of a map (or a floor plan) that was located at one end only (Figure 22, Figure 23, Figure 25) demonstrative pronouns *this*, *that* and adverbs *here*, *there* were in a frequent and effortless use. A test participant could explain something while pointing to a floor plan over a video link (Figure 22, left), and the other participants followed the line of pointing in the video stream:

- *Actually this is … the biggest room is the meeting room, this one is another office space, we have three more office spaces over here, and we have another big room and here shared bathroom and a kitchen. And we have men's and ladies' over here, so your idea of knocking out the partition and making a bigger room here is actually quite a good one, because then you can have three office spaces over here, meeting room and another office space over here, so yeah,* (turning to look at the opposite site video view), *that is a good idea.*

It was easy to establish from the video data that the test users effortlessly located/identified objects of deictic reference, for instance:

- *I don't know where you would like your office to be so can you just indicate where you want your office to be!*

A participant from the opposite site pointed to the floor plan:

- *Okay, I would like my office to be here, right here…*
- *That is my room!!!* (Immediate protest by a local site participant)

Figure 23 Across-site negotiation in progress
A participant indicates something with the help of a remote site map. Other participants follow the path of pointing (green arrow) in the video stream.

Attention was paid also to the addressees' paralinguistic / gestural response in concert with the speaker's pointing gesture. Such confirmative feedback from the remote site to the speaker implies that the addressees (think they) have 'got the point'. In such respect, the participants' verbal-gestural communication was in a great resemblance to a collocated communication - as can be seen from the following extract:

- *OK, well, Sam, you start off - do you know […], this main building here?*
- *Yeah-yeah-yeah.*
- *OK, you come up with the main gates, and if you want, I can meet you up there, because it is only five minutes' walk up there, but if you want to get there by yourself, walk slightly down the hill* (indicating the direction on the map), *and there is the first road on your left* (pointing), *when you cross over* (pointing, and emphasizing it with a zigzag movement; Sam nodding), *and if you just walk up that road, there will be several cars parked up there* (Sam nodding), *and you can see the […] sign, it is the pub just on the left…. this building.*

Coordinating contributions → effective communication: In a joint task, a remote participant had to give operational instructions to a local participant who performed accordingly in order for a joint task to be completed. The former could, for instance, indicate a suggestion that s/he wished the local participant to mark in the floor plan positioned behind the latter. The layout of the settings made the situation complex for the local participant. However, the participants co-operated successfully: the local participant looked in turns to the screen and to the floor plan (e.g. Figure 24) to see which point the speaker was indicating, then positioned the pen and again checked before marking that s/he had got the point right. The remote participant did not need to provide long procedural explanations; only should s/he wish it to be exactly 'spot on', s/he might add:

- *A bit lower … yet a little bit lower.*

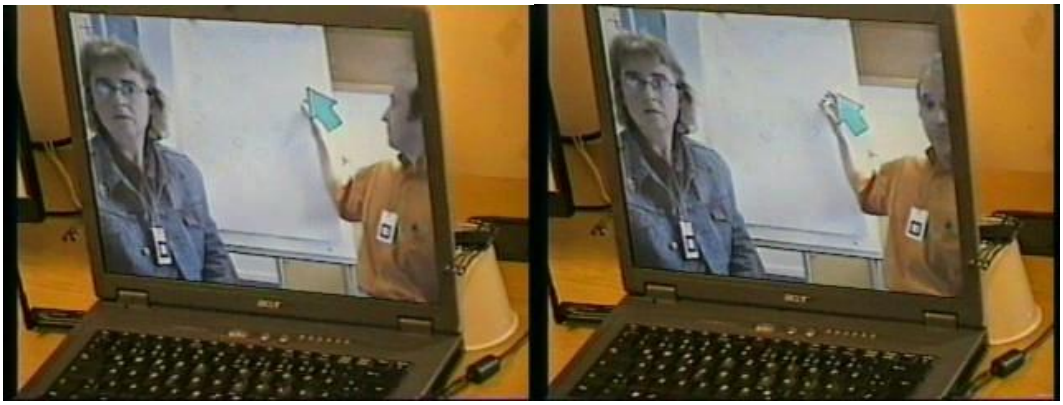


Figure 24 Alignment of actions in the across-site collaboration
A remote site person indicates particular position (the green arrow), and the local site person marks it in the floor plan.

Findings summarized: The findings from the video analysis show that the test users' (novel type of) across-site pointing was in every type of test task effective: introducing / addressing people, referring to remote objects, taking questions from the remote site, negotiating with the help of a hard copy, guiding directions and giving operational instructions to the remote site.

Also the narrative data from the post-test discussion and questionnaire confirmed that test participants found it effortless to establish joint attention to / shared understanding of the object of reference.

- *It makes our conversation more clear and easy to understand.*
- *Draws attention.*
- *Easy to show the things we are talking about.*
- *Get a clear perspective of what the others are talking about.*

The main findings concerning *effective gestural deixis* can therefore be stated as follows:

1. The test participants were able in the test settings *to guide* one another's *attention* to the object of gestural reference. If the speaker gave instructions by pointing, the addressee's *gaze turned to* the indicated point (e.g. Figure 22, Figure 23).

2. The test participants were able able *to locate (>to identify)* both the deictic origo (=person pointing) and the object of reference in their communicative situation without long procedural explanations. In other words, they were *able re-integrate the referential triangle* when a remote participant indicated something in the across-site direction.

3. As a result, they were able to establish *a shared understanding* of the object of gestural reference, and thereby, they were able *to coordinate their contributions* in a coherent manner in order to carry out co-operative tasks.

From the spatial design point of view, the test users were able to point in the test settings to objects at the remote site *without their pointing to lose its communicative function.*

Conclusion 1: In the light of the findings from the video data and anecdotal evidence, the conclusion is therefore that the hypothetical assumption that guided the design of the experimental video-mediated settings (p. 84) was on the right track: the outcome of design (=the test users' *interaction space*) *provided functional affordances for effective gestural communication*.

6.2 Adoption and Application for Communication

Assumption: The assumption concerning *usability* of the experimental test settings / interaction space was that the test participants would start using mediated pointing for their communication. The rationale was that once they discover a simple way to indicate remote objects, they go for it instead of relying on long procedural utterances.

The pilot test of Test bed 1[109] showed that a few minutes' familiarisation was too short for a novice participant to use all pointing modes that were mapped on the pointing device: should s/he focus on its operation, his/her task performance would suffer *(>cognitive overload).*

Therefore, the test procedure was modified and the application engineer played the role of a chairperson in the first set of tests. Consequently, the analysis then focused on the participants' responses to pointing. Being familiar with the application, the application engineer was able to use the pointing functionality to its full potential and thereby, to provide best practices for the other participants. It was also assumed that the person next to would exploit his/her favourable position to learn a new pointing strategy and at some point of the test, try to use it.

Test bed 2 featured one pointing mode only. In the tests for two-way pointing, there was one pointing device available at both sites. Consequently, all test participants had a possibility to use pointing functionality. It was possible to indicate both local and remote objects in a mediated way.

[109] Test bed 1 implemented by Guido Kuehn, Test bed 2 by Swen Walkowski.

Method and outcome of analysis: The following questions guided the analysis of the video data:

- Do the test participants use pointing functionality?
    - What for do they use it?
    - How do they use it?

In order to give the test participants' experiences a user's voice, narrative data from post-test discussions and their open comments (Appendices 5 & 6) was thematized (Table 12):

- What kind of perceptions did the test users have concerning the pointing application / interaction space?
    - Was it helpful or not?
    - Was it convenient or not?
    - Was it easy to use or not?
    - What problems did they report?
    - What benefits / shortages did they report?
    - What amendments did they suggest?

Frequent use: The video data from the test sessions clearly shows that the test users started using the pointing functionality (Figure 22, 23, 25, 27, 28 and 30). In the tests for one-way pointing, only the person next to the chair had a possibility for mediated pointing whereas in the tests of two-way pointing, it was possible for all participants.)

In different test tasks, pointing functionality was in a frequent use among test users. Even competitive situations were recorded where several participants tried to get hold of a pointing device (Figure 30). It is much likely that the (then brand-new) pointing device had a motivational role, too.

The findings from the video data suggest that pointing functionality saved time and effort as it was simple just to show things instead of giving long verbose descriptions.

Use for multiple purposes: The test participants were using pointing in multiple tasks for different purposes. From video data it could be seen that a typical case was to indicate the position of a remote object (deictic reference). Pointing had in many cases also an iconic nature: the size of an object, the extent of an area, to describe a path, a direction and a connection between different components.

Apart from indicating spatial properties and relations, the participants used the video-mediated pointing in a paralinguistic way: to emphasize something they were just explaining, with a zigzag / spinning movement.

Taking into consideration that the participants were carrying out a joint task, cases of disruptive use were recorded, too: it is likely that the novelty of the pointing device excited some participants to the point they could not resist playing with it.

Preference of mediated pointing: The participant were using mediated pointing not only for the remote (Figure 23) but also for the local objects (Figure 22, left, Figure 25). The likely reason, judging from the video data (e.g. Figure 22 left), is that they found it more ergonomic and convenient to point to the representation of a local object via the screen than to turn backwards to show it.

There was also a case when more participants wished to point at the same time than there were pointing devices available (Figure 30): Novelty of the device clearly played a motivating role when they tried to get hold of the device.

Drawing on the video data (Figure 27), people seem go for the least effort option: once they discover there is a simple way to communicate in the test settings, they go for it, and may even favour it. It is, however, not clear whether such a preference is long-lasting or whether it fades away with the novelty impact of the gadget.

Helpfulness and convenience in communication and collaboration: The data collected from the post-test discussions suggests that the test participants found mediated pointing, in many respects, beneficial, helpful and convenient:

- *It makes really easier to pass information from one to the others.*
- *Turn taking feels completely normal.*
- *Less crosstalk, obvious when it is your turn to talk.*
- *Great co-ordination*
- *Task-oriented work gets clearly easier.*
- *It is good for remote meetings, wastes less time.*

The settings supported test users in catching the attention of an addressee, in directing their attention to the object of reference, and in turn taking. Mediated pointing also reduced their possibility of misunderstanding, supporting them in keeping track of the ongoing discourse. The settings supported task-orientation and carrying out joint tasks both in a technical and psychological consideration: a ‘threat’ of the chair next pointing to you kept the participant at least vigilant. Furthermore, mediated pointing was considered efficient, which is a real advantage in a videoconference (Table 12).

Table 12 User comments on usability

ADVANTAGES	DISADVANTAGES
- Easy to show the things we are talking about. - Task-oriented work gets clearly easier. - It's good for remote meeting, wastes less time - Adds dimensions to flat videoconferencing - Being able to highlight the person who is talking among a crowd / Indicate things / point at items / highlight areas - Colour coding definite asset. - Different colour bubbles help in differentiating between members of different groups. - Camera feature is good as you can recall options easily. - Liked the tags – ability to make notes. - Good to give name – extremely useful.	- The remote control device is a little bit difficult to use. - Handling of the remote pointer. - Not really easy to control. - Sometimes AR is little [too?] sensitive.
EASE TO ORIENT / NAVIGATE	**DIFFICULTY TO ORIENT / NAVIGATE**
- Draws attention. - Easy to point something that should be noticed on the screen. - Easy to know who is being pointed [to]. - The difference for the better is that the object pointed to and the person talking should be identifiable at the same time so that you can see the object *and* the persons talking. - Arrows are also helpful for indicating directions.	- What confuses me about v c is the reverse image & movement of pointer takes skill & practice - Still very sensitive pointer. - The arrow is too large to point more accurately. - The pointing device is NOT user-friendly; you can't point to an item clearly and accurately. - Some time lag.
HELPFULNESS IN COMMUNICATION	**UNHELPFULNESS IN COMMUNICATION**
- It helps really to focus on, to specify what one didn't understand. - Communication is fluent. - Highlighting the focus of a discussion. - Great co-ordination. - Better interaction. - The possibility of misunderstanding reduces. - The communication speed increases. - Able to chair the meeting. - It helps to introduce each other from the remote sites to the other especially when you don't know the persons. - [Spinning] arrow made us all laugh, good for ice breaking, silly in serious meeting.	- Missing a natural feeling of meeting people. - May slow down meeting but advantages outweigh disadvantages. - The pointer distracted us from the main point of the task. - People may focus too much on AR instead of just using their hands. - Irritating if people use pointer when they are not talking or pointing to anything particular. - It may cause some misunderstanding if the pointer-user is not an expert. - Pointer should be used very carefully to keep it still.
CONVENIENCE	**INCONVENIENCE**
- Less crosstalk, obvious when it is your turn to talk. - It makes our conversation more clear and easy to understand. - Get a clear perspective of what the others are talking about. - Can clearly know which is the point. - Use of a pointing device is good. - Easy to use pointer. - Fun to use.	- Some people will not appreciate being 'put on the spot'. - It made one feel a little bit uncomfortable if you don't have an idea to the question … the bubble and pointer on you; it is a bit discomforting. - Slightly unnerving seeing pale circles etc. on your face. - Looks more artificial. - Nothing made it worse.

Unhelpfulness and inconvenience: Apart from positive comments, there were also opposite views. Among them, some participants found that the video-communication was lacking a natural feeling of meeting people. It also looked more artificial; for someone, the reverse image from the local site was confusing.

Ease of use vs. problems in use: The pointing device was sizeable enough be seen in the video image. As a result, it was easy for the participants to link the connection between the origo of pointing (=the pointing person at the remote site) and the object of pointing.

Using the pointing device (WiiRemote) was considered, in general, less natural than finger pointing. Many test participants found the pointing device easy to use, while some others experienced difficulties with using its keys. For some, the pointing device was a bit too big and therefore, they found difficult to reach different keys.

- *The remote control device is a little bit difficult to use.*
- *Not really easy to control.*

It became clear from the user comments that mapping different functionalities on the pointing device was not the most successful one: pointing should be, as a performance, simple, ergonomic and intuitive. For instance, turning on/off pointing was not as intuitive as it should have been (in Test bed 1), and in order for the user to feel when it was slipping out of the video frame, haptic feedback would have made a difference (in Test bed 2).

As problems concerning the very pointing device go beyond the focus of this thesis (*interaction space*), they are not discussed here in further detail; it is only stated that the test users provided several ideas (some of them mentioned in Table 12) for further amendment of the application.

None of the problems experienced by the test users with the pointing device hampered/prevented them from carrying out their joint test tasks; yet, it is evident that any cognitive / ergonomic problems had at least some impact on their performance.

Conclusion 2: It is concluded here in the light of the user feedback that several immaturities of the pointing application call for further development, in particular the mapping of the functionalities on the keys of the pointing device.

However, none of the criticism by the test users calls to question the design principles that guided the design of interaction space: both the video data and the users' comments confirm that *the interaction space of the experimental test settings supported the participants' coordination of contributions.* Specifically, it supported their

1. individual / joint navigation,
2. mutual referencing,
3. turn taking,
4. keeping track of the ongoing focus of communication, and
5. (by reducing need for procedural verbalisation) more natural multimodal communication and interaction.

6.3 Adaptation to Unfamiliar Context of Interpretation

Assumption: According to the initial assumption, the test users were expected to adapt in the experimental test settings and to be able to navigate it.

Yet, it was assumed highly possible that some misinterpretations would occur in the initial use, in particular, that somebody would forget the constraints of a videoconference and point with a finger to a remote site object.

The users were encouraged in the test session to think aloud so that their initial thoughts and impressions could be captured *in situ*.

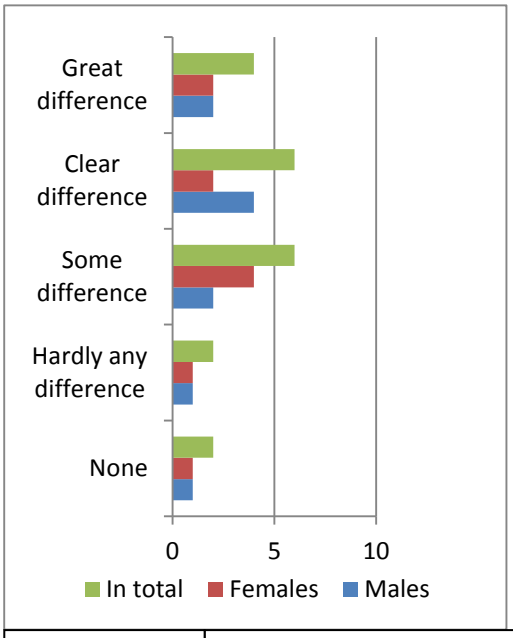
Method and outcome of analysis: In order to find out whether people really adapted to the settings, video data from the test sessions and narrative data from the post-test discussions were analysed: attention was paid to possible misinterpretations of the spatial frame and to any behavioural signs indicating that adaptation was taking place.

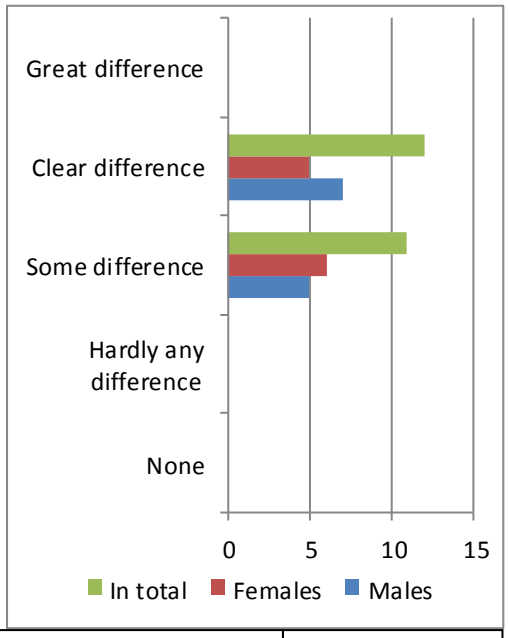
In addition, the test participants were asked to rate their perceptions of Augmented Reality (AR) -enhanced video-conferencing in comparison to ordinary videoconferencing using Likert scale no difference (1) - great difference (5).

Perception of AR in videoconferencing**:** The views of Test bed 1 users concerning the difference between traditional vs. AR-enhanced videoconferencing varied from no difference to great difference; yet the majority of them described it as *clear* or *some difference* (Table 13). Test bed 2 users considered there was *clear difference* or *some difference*. In both user groups, the male users regarded the difference to be a bit bigger that the female users did.

Yet, the interaction space was in principle similar in both versions of the test bed. The main difference between them was that Test bed 1 was for one-way pointing whereas Test bed 2 gave the possibility for *two-way pointing*. In the latter case, the participants had better possibility for across-site interaction. That is likely to explain the difference between the views of Test bed 1 users and those of Test bed 2 users.

Table 13 Perceptions of Augmented Reality (AR) in videoconferencing





Test bed 1	Does AR make any difference in videoconferencing?	Test bed 2

Initial orientation**:** Judging from the facial expressions and initial comments by the test participants, they had, indeed, ended up in an unfamiliar situation. Yet, when a test participant was for the first time pointed to from the remote site, s/he instantly recognized that attention focused on him/her. The most poignant case was when the chairperson pointed to someone at the remote site: the latter looked perplexed[110]; her gaze went zigzag in disbelief, she pointed to herself with a finger and exclaimed, before taking the floor:

- *Me???*

She could clearly not grasp how it was possible. The test users were initially puzzling and trying to figure out how the pointing mechanism worked. Their behaviour reflected, in general, keen interest and inquisitiveness on one hand, and on the other, cautious attentiveness, at times even signs of awe. Similar affect-laden facial expressions can be spotted when small children try to figure out the workings of something they are not yet able to understand. Exclamations could be heard:

- *Wow, how did you do that?*

[110] No illustration because of consent declined.

Figure 25 Adaptation to the frame of communication
The participant indicates a remote site object asking what it is (>left display), and later on, explains a local site object by indicating it on the local site view (>right display)

Context model adjusted/updated: In the initial phase, the user's cognitive expectations did occasionally fail, as expected: the test user had to go out of the familiar frame of communication/interpretation/action, and there was a particular threshold to be crossed. That could be seen from the following episode in the video recordings: in the middle of a negotiation, a test participant pointed with a finger to a floor plan that was physically available at the remote site. At the same, she realised that her gesture was ineffective in guiding the remote participants' attention, and she burst out:

- *Oh gosh, I cannot show it!*

Being immersed in a joint task, she took the hybrid and augmented context for a real space. Another cognitive clash followed, when the chairperson passed the pointing device to her, urging:

- *Yes, you can with this!*

She seized the pointing device starting to drag it on the desk as if it were a mouse. When the chairperson showed how to use it, she was within 10 seconds indicating remote objects in a fluent and confident manner.

The above revealed a clash between *the user's mental context model* and *the real situational context*. It also showed that the person profited from her position next to an experienced user, and she exploited an innovative pointing strategy as soon as she discovered it.

Immersion: The video data also revealed that only a couple of minutes later, the same participant grabbed the pointing device from the chairperson's front wanting to make changes to the floor plan at the remote site (Figure 28, left). It looked like the novel pointing strategy was already a routine to her. When the video clip was later shown to her, she confirmed that she had no recollection of grabbing the pointing device: she must have been so focused on a joint task that she hardly paid any attention to it.

In her case, using the pointing application appeared both easy-to-learn and easy-to-use. Yet, that was not the case with all test participants as can be seen from the comments concerning usability of the pointing device (Table 12).

Contextualising spatial performance: The video recordings from the test sessions revealed how test users flexibly modified and adjusted their communicative behaviour according to their situational context. It could be established from the video data of the tests sessions on Test bed 2 (where the camera was between the two displays) that the participant's gaze direction monitored the situation switching between the local and

remote-site view (Figure 28, right; Figure 25). To take on example, an addressee was following the path of remote pointing on a local-site screen while regularly glancing at the remote-site view for social cues (the remote person's body language).

Similarly, a local speaker glanced at the remote site for a visual feedback (>gestural actions/ responses there). Such communicative strategy supports the participants' *coordination of perspectives* and their *turn taking* in a conversation.

In navigational terms, the local site view (reverse mirror image) was for someone, however, confusing:

- *What confuses me about v c is the reverse image…*

In the video data, there was also an episode when the pretty sizeable arrow representation crossed the video view, and a person moved out of its way as if shunning a falling object. One possible interpretation is that the participant took the pointing representation for a real thing; another interpretation takes a more social-communicative point of view. That is clearly the case in another situation (Figure 30) where a test participant turns out of the way so that a remote-site participant can show something on the floor plan.

Timing and duration: The video data showed that if a test participant was pointed to, s/he interpreted it as a call to respond (Figure 21).

- *Easy to know who is being pointed [to].*

If the pointer representation lasted on one's face persistently, people started to feel uncomfortable - as if being stared at, or, compelled to talk. From the communication point of view, it was found that the duration of a pointing representation is critical: it has last long enough in order to capture the attention of the addressee, but no longer; otherwise it becomes counteractive.

Pointing representation: Judging from the user comments, no one had complaints concerning the salience and the shape of a pointing representation whereas both the size and the colour were criticized. The blue colour of the representation for a human object (in Test bed 2) did not appeal to participants who did not wish to look ghostlike, as one of them put it.

The size of the arrow was found good for indicating directions whereas for pointing to small objects, it was too big.

- *Arrows are also helpful for indicating directions.*
- *The arrow is too large to point more accurately.*

Figure 26 Hand pointing to a local object

Figure 27 Mediated pointing to a local object
A participant (left) is pointing to the local site object over a video link (even though the map indicated is just behind him). The person next to him follows by gaze the movement of an arrow.

Figure 28 Quick adaptation to a novel frame of interpretation
The person gesticulating (left frame) is seen just grabbing the pointing device (without even noticing it) though she had only a couple of minutes earlier used it for the first time.
When a remote participant was pointing to the object (right frame), people followed the object of reference (arrow) from the local site view and sought, and at times, glanced at the remote site view for procedural information.

Both the size and the colour have to do with salience. As the pointing representations clearly caught the participant's attention, they were salient (perceptible); that again is a functional precondition for guiding the attention in a video-mediated way to an object of reference.

Colour coding was considered by the test users to support communication (navigation), and it was also suggested that even the pointing device be colour coded according to the representation.

Pointing experience: The users' experiences of being in the focus of pointing varied from person to person. Some participants appeared to enjoy being in the centre of attention (Figure 21, right), whereas some others reported uneasy feelings, when a pointing representation for the first time appeared on their video image.

- *Some people will not appreciate being 'put on the spot'.*

Some participants found the video-mediated pointing was a bit unnerving; at the same, it made them more focused on the joint agenda:

- *It made one feel a little bit uncomfortable if you don't have an idea to the question … the bubble and pointer on you; it is a bit discomforting.*

For someone, being pointed to gave a feeling that he was important. Some participants became self-conscious of their visual (media) appearances, and some others again found their co-participants' dealings distracting in the midst of a joint task:

- *Irritating if people use pointer when they are not talking or pointing to anything particular.*

Code of social conduct: At the end of a test session, there was a coincidence that made apparent the role of the code of social conduct in the video-mediated settings. While talking sideways to a person, the chairperson was holding the pointing device, and it was coincidentally pointing in the direction of the remote site display. The pointer representation then happened to focus on the private parts of a remote participant causing an awkward social situation at the far end.

Video-mediated communication is governed by the same expectations of considerate social conduct as is collocated communication. Due to a more complex context of action and interpretation, however, unintentional breach of conduct is more likely to occur. Therefore, context-appropriate social conduct demands particular attention in the video-mediated settings.

Figure 29 Across-site collaboration in progress

Figure 30 Competitive situation
(On Test bed 2)
Everyone would like to show their preferences: two arrows can be seen on the screen, and furthermore, one person is pointing with a finger. A participant is simply snatching the pointing device from his teammate; there can be seen a bit of hand wrestling before the latter gives in.
Face detection can be seen in function (bottom frame).

Conclusion 3: Drawing on the analysis of the *in-situ* observation and the video recordings of the test sessions, the test users' adaptation to the context was, apart from (initially) affect-laden, astonishingly quick.

There was clearly a cognitive threshold to pass and related affect-responses, but soon the participants were able to make use of their *situational frame* of action/interpretation in the search of relevant information to update *common ground* in their across-site co-operation.

In perceptual terms, the pointing representation was salient enough as an attention getter, and an Augmented Reality -visualisation easy-to-interpret in the video stream.

The test participants adapted to the designed interaction space (=hybrid/augmented frame of interaction) and effortlessly navigated it, despite several immaturities that were pointed out concerning the mapping of functionalities on the pointing device.

6.4 Outcome of Case study 1

The experimental test settings provided *a visual perceptual common ground* so that the test participants could see one another, introduce objects to the attention of one another, establish a joint attention / a shared understanding, and thereby, deliver a coherent joint line of contributions. As a result, they were able to collaborate in a seamless way carrying out co-operative tasks over a video link in a successful way (e.g. Figure 29, Figure 30).

To put it in design terms, the settings enabled *the coordination / reciprocity of perspectives* by two displays at each site, one representing the local site view, the other one the remote site view. The test bed (Kuehn et al, 2007) was capable of tracking the direction of pointing, and visualising the indicated point in the video stream.

As can be established from the findings in Chapter 6.1-6.3, the outcome of the user tests supports the assumption of an effective deixis: the communicative function of a pointing gesture is inferred, at the bottom, from *a spatial triangulation* of

1. the speaker's origo of pointing,
2. the position of the object indicated, and
3. the addressee's origo of observation.

The participants have to perceive *who is pointing to what.* Yet, a shared information space *as such* does not guide the attention to a particular point (*focus*). The findings show that in order for an interaction space to enable the communicative function of an across-site pointing gesture, it has to meet two other conditions:

1. Both the deictic origo (=who is pointing) and the deictic referent (=what is pointed to) have to be *within the scope of a shared information space* (>interaction space) (Figure 9).
2. The connection between the deictic origo (the person pointing) and the deictic referent (object pointed to) has to be made *perceptible* from the addressee's point of view; only then the settings enable *a referential triangle* for the participants to establish *a joint visual attention on an object*, and thereby, *a shared understanding of a deictic referent*.

The mechanism of (gestural) deictic reference could therefore be translated into design relevant terms as follows:

In a communicative situation, *shared understanding* of an object of reference relies on *deictic reference*.

- Deictic reference technically relies on joint attention.
    - Joint attention relies on mutual attention and on an object of reference in a shared information space (=an object presented to the attention of the addressee).
        - Mutual attention relies on mutual awareness.
            - Mutual awareness relies on shared information space.
                - Shared information space is a physical connection between individual perspectives.

(See Figure 9, Figure 10 and Table 4)

6.5 Coordination of Complementary Perspectives in Case Study 2

Multimodal deixis had in Case study 1 turned out to be a helpful tool in the investigation of the role of space in video-mediated communication and interaction. A particular bonus was that when situated, deixis encompasses all the critical dimensions of communication and interaction that have to be taken into account in the spatial design: the spatial, the social, the modal and the cognitive one. (See Figure 9, Figure 10, Table 4, Chapter 5.5.1).

In Case study 2, complementary roles of communication/collaboration were regarded as complementary spatial perspectives. In order to enable reciprocity of perspectives, the role perspectives had to be technically orchestrated to provide reciprocity of perspectives in audio/video modality.

Applying situated deixis in the design helped to reconciliate between two different perspectives, that is, the cognitive one that is necessary from the navigation point of view, and the ethnographic one that is required in order to capture social-spatial practices as patterns.

The cognitive approach helped to lay out the human-environment interface in spatial design-relevant terms:

- as an information channel between the individual points of observation/action to provide functional mutuality in the audial/visual modality, and thereby, to enable mutual awareness;
- as a triangulation of the point of observation, the deictic centre of pointing and the object of gestural reference in order to provide effective gestural deixis.

An ethnographic observation of lecturing, and discussions with the observed teachers made possible to describe design-relevant patterns in the teacher-student interaction. Such regularities develop in cultural practices over time and they become common *social-cognitive* choreographies that individual people tend to follow in their encounters (>spatial pattern). Thereby, they cover a wide range of individual differences in human communication.

Assumption: Functional affordances for shared understanding were assumed to result from the coordination of the complementary role perspectives (in a topic relevant and task sensitive way). Therefore, situated deixis was in Case study 2 assumed to be an ideal tool in the practical design case, that is, in the coordination of complementary role perspectives (→*reciprocity of perspective*) for synchronous multi-site communication in a traditional lecture theatre (=site) for Biomedical Sciences (=field) that was due to be converted for video-mediated lecturing (=activity) across three universities.

Methods: The deixis-based design principles were applied in the practical design case. First, the key role perspectives were specified (=teacher, student). Thereafter, initial orientation to the design task took place through a role play in the lecture theatre where the constraints of the venue (Figure 19) were mapped and key role perspectives were considered (Chapter 5.5.2)

The specific role requirements were mapped through observation of a set of lectures in the subject field (**Figure 20**), communication with the stakeholders in focus group discussions and one-to-one conversations with the observed teachers (Chapter 5.5.3, Appendix 7).

Attending a number of lectures given by different teachers at three universities gave insights in lecturing practice in the subject field. It also helped to map differences between individual teachers in their communication and interaction with the students (use of space, delivery, range of methods, strategies and resources used).

The role specifications then guided the spatial arrangements in the lecture theatre (Chapter 5.5.4). After implementation, preliminary feedback from initial use (showcase) was collected (Chapter 5.5.6, Table 9) paying attention to the following questions:

- Does the teacher see the whole audience without extra effort?
- Are the students in an (as) equal position (as possible) in terms of
    - the focus of communication from the local / remote site?
    - the teacher's attention and procedural cues from the local / remote site?
    - interaction during question and answer (Q&A) -session?

6.5.1 Lecturing as Spatial Performance

The following patterns were mapped based on the lecture observations. They were considered particularly relevant to the teacher's role perspective in lecturing:

1. Positioning: The observed teachers had an individual 'choreography': either they stayed at the lectern (console) or moved between different positions while delivering a lecture. The position was then dependent on the type of activity: they might move a few steps back and forth in their 'base', go to another position to show something, and yet, to another place to interact with the students. Some had a habit of walking circles in front of the class while talking[111].
2. The three critical reference points (Figure 31):
    - the teacher's notes
    - the focus of the topic indicated on the screen/whiteboard
    - the gaze contact with the audience to adjust the speech delivery ('audience design'), and a turn taking properly.'
3. A prominent and meticulous pointing: (The teachers did make sure that the students get the object of reference right. Furthermore, pointing had a clear function in floor control (when the teacher was taking questions, or, someone among the audience wished to ask a question). Apart from pointing, a whole range of iconic gestures was

[111] Assumingly, it might help them to pace their speech delivery, to balance tension in the focus of the audience's attention, and thereby, to better focus their thoughts.

used to illustrate the shape, size, direction and process. A hand gesture typically also emphasized a verbal utterance.

4. The two communication modes: The principal modes of communication could be described as *a presenation mode* (Figure 32) and *an interaction mode* (Figure 33) when the teachers switched from delivering the contents to taking questions. Their frequency to monitor how well their communication 'sank in', varied from person to person. For someone, asking questions was a frequent confirmation (and exploitation of we-intention) along the presentation:

   - *Are we all right so far?*
   - *Everybody up to this point?*
   - *Any questions at this point?*

   Someone else had a tendency to challenge the audience: a quick round of questions was a way to train the students in reasoning and argumentation:

   - *And how might I know that? … Think about it!*
   - *I just explained it. … Now: what is the answer?*

   - *What do you think it might mean?*
   - *How might that happen?*

5. Resources: A typical resource used were PowerPoint slides (occasionally embedded slides with animations). A flip chart and/or the white board were used for additional information. In addition, tacit artefacts and the document camera were employed.

6. Controls: Some teachers mentioned that they would prefer just to walk in, to log on, and to retrieve their resources online. The use of controls varied from person to person: someone went to the laptop to change a slide; someone else used a remote control. Also the lectern console and the smart board were used.

Figure 31 The three critical reference points
The notes, the slide on the screen, and the audience are the key reference point from the teacher's perspective. Visual contact with the audience helps to guide the attention of the students to the focus of the topic (on the slide). It also helps to assess whether the audience is following, and whether they (appear to) have understood the message.

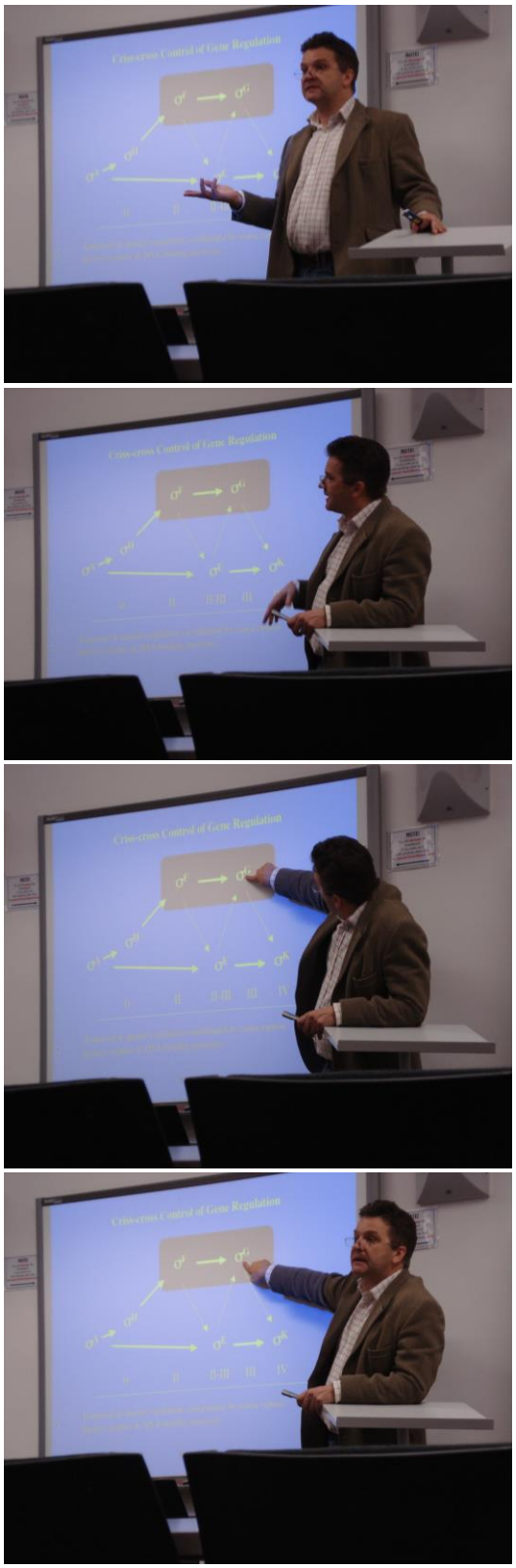


Figure 32 The teacher in the 'presentation mode'
Pointing gesture and gaze direction provide cues and guide the attention of the audience to the object of reference. A gaze contact with the audience both initiates and ends the guiding.

Figure 33 Teacher in the 'interaction mode'
A different pattern of communication appears when the teacher wants to check whether the audience has understood what has been explained: he focuses on the student, probing and encouraging a reply also in a gestural way.

6.5.2 Specification of Role Perspectives

The teacher's perspective:

1. From an ergonomic and communicative point of view, the teacher has to see the *whole* audience with a single glance in order to monitor nonverbal cues and to modify his/her speech delivery compatible with its reception (Figure 31).
2. It should be possible to use multiple resources in a video-mediated lecture.
3. There has to be an effortless mode switch when the teacher takes questions (Figure 33) from the (whole) audience.
4. The layout of the input/output devices (cameras, microphones), and the positioning of the controls have to be ergonomic.
5. Operating the controls has to be as simple as possible so that the teacher's focus on communicating the lecture's content to the students is not disrupted. As the lectures are packed with information, there is no time for complicated technical maneuvering.

The local student's perspective:

1. The local students have to see clearly *what* the lecturer is explaining. In addition, they pick up the lecturer's nonverbal procedural cues (**Figure 20**).
2. During the Q&A sessions, availability of the microphone is also important in order for them to contribute to an across-site conversation.
3. They might also wish to see who is in voice at the remote site during Q&A sessions.

The remote student's perspective:

1. From a motivational point of view, the remote audience has two more requirements in addition to the above: the camera view on the lecturer has to be from a particular angle so that the remote student has an impression as if s/he were among the local audience: the teacher is talking to *us* instead of talking to *them.*

2. As teachers move a lot while lecturing[112], a (high quality) camera tracking would be a technical way to enable more freedom for the teacher to follow his/her habits. Yet, a simple way to go around the problem is to pre-set the camera view(s) according to each teacher's spatial patterns so that s/he is not going out of the remote audience's sight; otherwise, they might lose important procedural cues.

6.5.3 Implementation

Positioning a display of the remote site view close to the teacher's eye level would have meant sacrificing some of the seating capacity of the lecture theatre. As that was not feasible, two large screens were fitted in the back of the lecture theatre, one for the remote site view(s), the other for the slides.

The outcome provides the same advantage: the teacher can see the multi-site audience with a glance and gather from the participants' nonverbal cues (visual feedback) whether or not they follow along what is being explained. In the other display, s/he can effortlessly see the slide that the students see on the front screen. Such spatial arrangement is ergonomic because the teacher does not turn forth and back; at the same, it supports the teacher-student interaction because s/he can easily spot should any of them – local or remote – wish to ask a question (Figure 34, above).

A wide screen for the audience in the front of the class, and a console with three beamers above (Figure 34 below) make it possible to display multiple views side by side. For instance, one way to increase a sense of a social presence during the Q&A sessions would be to display relevant views from the remote site(s)[113].

The camera focusing on the teacher is positioned between the two back screens (Figure 34). As a result, when the lecturer takes a gaze contact with the local audience, the *whole* audience would be addressed. When taking questions from and giving the floor to a remote student, the teacher watches the left back display, quite next to the camera focusing on him/herself. As a result, the impression at the remote site will be as if the teacher were looking at people sitting there.

[112] It would require a lapel microphone, too. In this presentation, attention is mainly paid to *gestural* deixis; therefore, the focus is on *visual* information.

[113] The control of several views requires technical assistance by AV staff.

Figure 34 The teacher needs visual feed-back to design his/her talk properly
Two screens were therefore at the back wall of the lecture theatre so that they provide the lecturer a view on the remote audience (left) and the view that the students see on the front screen (above right)
The camera in the middle of the two screens at the back wall provides the remote viewers the feeling as if the lecturer were looking there when s/he looks either at the local audience or at the screens.
During the Q&A sessions, a wide screen in the front of the class makes possible to increase the feeling of social presence and inclusion by providing a view of people communicating from the remote site(s).

6.5.4 Initial User Feedback

Technical aspects of the showcase session such as quality of transmission, and communicative aspects such fluency of Q&A session, were rated by the initial users according the Likert scale 1…5 (Table 10, Table 14).

Overall impression: The user feedback indicated that the settings met the expectations in many respects - in some aspects exceeding them - but left space for specific concerns in technical robustness and reliability (Table 14, bottom tables).

The respondents saw potential and advantages of large-scale videoconferencing in the educational use:

- *I found the lectures were nearly the same as lectures in traditional co-located settings.*
- *Ability to converse between sites.*

Showcase lectures (5): The feedback of the lectures was positive: except one, their ratings fell into category 4 or 5 on the scale difficult to follow (1) ... easy to follow (5). Due to a local pixilation problem, one remote presentation was temporarily hampered, and in another case, the lighting of the lecture was unfavourable at the remote site so that his gaze direction and facial expressions could not be properly discerned. The comments about the presentations were positive.

- *The lecturers made more of an effort to be clear than they might to normally, which is good.*

Camera view to guide attention: It is important that the camera focuses on where the audience's attention should be, that is, guiding the attention to deictic reference.

- *We could see well the remote audiences at […] and […] but there was no opportunity to see close ups of people (for example when asking specific questions).*

Not only the focus but also the scope is important: it gives the visual context for the interpretation. In the subject field discussed, the teachers rely heavily on pointing gestures in making sure that people get the referents right. The remote students need to see, in order to cover the relevant context, not just the teacher's face but also his/her hands. Discerning gaze direction and facial expressions of the lecturer in relation to the object of pointing is important as such cues support the interpretation of the message.

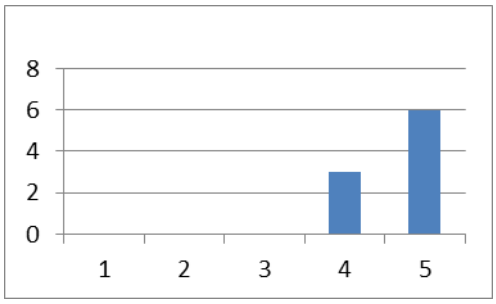
In case of delay in video transmission, the lecturer needs see the video image of him/herself transmitted, because it helps him/her to pace his communication according to the overall context.

Across-site conversation / interaction: The chairing of the showcase met the expectations of the audience successfully:

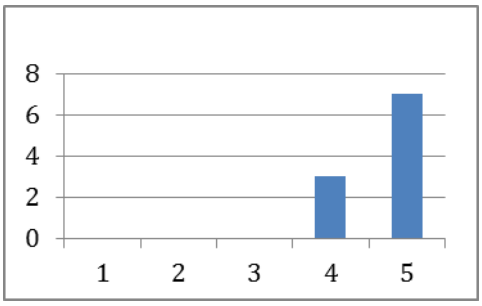
- *...very good were [...] chairmanship and the helpful role of the moderators at each side.*

Quite collocated spontaneity and fluency cannot be expected in the video-mediated communication as can be seen from the following comments; it will take some time for the teachers and for the students to get used to a more complicated context and to the practical manoeuvring of Q&A sessions.

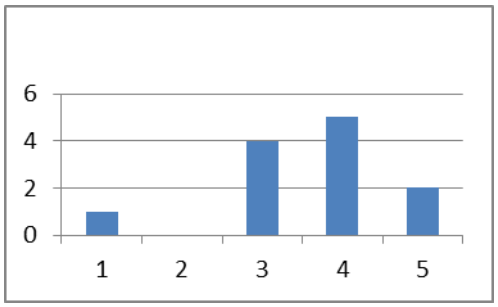
Table 14 User feedback from showcase



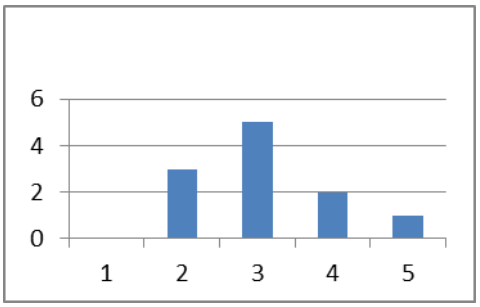
Presentation A (n=9)
Difficult to follow 1 ... easy to follow 5



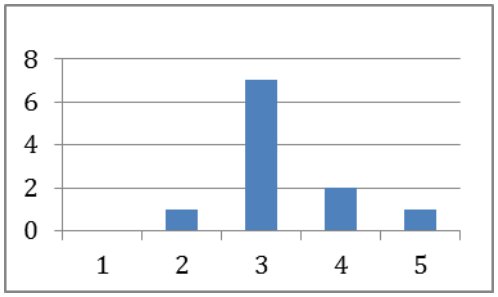
Presentation E (n=10)
Difficult to follow 1 ... easy to follow 5



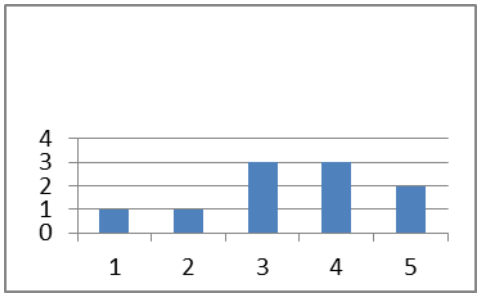
Camera view (n=12)
=direction & focus to guide the attention
Below satisfactory 1 ... very satisfactory 5



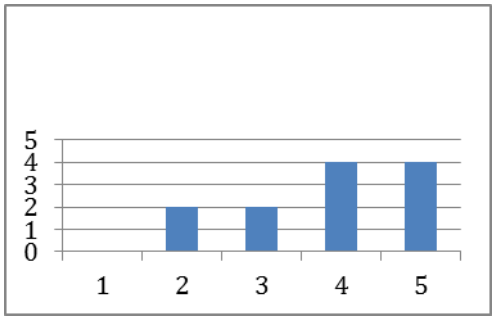
Sense of presence of the remote audience (n=11)
Below satisfactory 1 ... very satisfactory 5



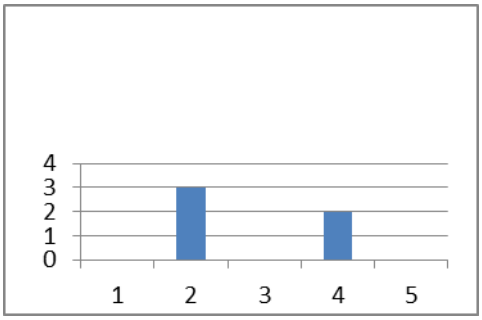
Fluency of Q&A sessions (n=11)
Below satisfactory 1...very satisfactory 5



Other people involved in conversation (n=9)
Difficult to follow 1 ... easy to follow 5



Technical interruptions (n=12)
Below satisfactory 1 ... very satisfactory 5



Technical disturbances (blurring, pixilation) (n=5)
Highly disturbing 1... not at all disturbing 5

- *Not very vibrant Q&A.*
- *Fewer questions from students.*
- *I felt it lacked the sense of occasion of a live lecture.*

During the across-site conversations, the audience(s) has/have to hear clearly what the question is. If possible, they tend to look at the person who is in voice. Local moderators are therefore needed at each site to assist in managing practicalities in the across/site conversation. If the co-operation of the chair, local moderators and AV staff is seamless, the turn taking is fluent, and each audience gets situation-relevant camera views on their screens.

Challenging context: It became clear from the teachers' comments that the hybrid settings are, due to their complex spatial nature, a challenging frame for lecturing:

- *As a one-off event it was a novelty to see the live three-way connection.*
- *Lectures would have to be planned carefully in advance.*
- *Seems to demand more concentration.*
- *There would be less scope for spontaneity, i.e. drawing additional diagrams on the spot, or modifying the content of one lecture to adapt to changes in the previous one.*

Technical quality: Video and sound transmission met in general the audience's expectations. At one site, a quite temporary pixilation occurred[114].

Reliability of the system was a source of worry: as the study modules are packed with information, a technical failure would jeopardise a whole lecture. Teachers were cautious in their optimism, expressing the following kind of concerns:

- *I left concerned that this system is not sufficiently robust to enable the delivering series of lectures without breaking down at times.*
- *[....] the slide corruption and the worry that this would engender as to whether this was going to be a frequent occurrence in timetabled teaching.*
- *The quality of sound and video from remote sites was satisfactory but the a-synchronicity of sound and image is rather unsettling and remains an issue.*

As it is likely that technical problems occur at times during videoconferencing lectures, each session should to be available to the students in the recorded form. Also a contingency protocol is required between AV staff, lecturers and moderators. Knowing *how* to cope with any local and across-site occurring in sound connection, video connection, data transmission, and in accessing (online) resources, saves people at least from extra problems. It has to be clear for everyone what they need to do. The best way would be to arrange a simulation where the following points are specifically clarified:

[114] As technical configuration, lighting, and soundscape go out of the topic of this thesis, they are left out of further discussion.

Figure 35 Across-site conversation in progress
The positions of the camera in the back of the lecture theatre focusing on the speaker and the display projecting the remote view enable a sense of a mutual eye contact.

- What kind of information is relevant the local audience?
- What kind of information is relevant to the other site(s)?
- Who informs them?
- What are the ways to go around that particular type of problem?

Collecting feedback from the problems met is not only needed for further design, but it serves the whole life cycle of the product/service. It became obvious in the course of the case study that more communication between the teaching staff and AV staff is needed: recording the problems met and of ways of solving / avoiding them, would benefit the users themselves, the maintenance of the system, and the new members of staff. Sharing experiences between different user perspectives would thereby contribute to the improvement of the novel learning interface.

6.5.5 Observations from Emerging Practice

Judging from observation of curricular lecturing in the midst of the collocated students, the sightlines for mutual conversation between the teacher and a remote student were natural-resembling: when a remote student asked a question, the experience was as if the person were one of the back benchers in the same lecture theatre. The teacher was then focusing on the left screen behind the collocated audience, and the camera transmitting video image of him to the remote site was just next to it. The situation there must have been as if the lecturer were looking the person into the eye (Figure 35).

Based on the observation of emerging practice, the teachers adapt to their novel context of lecturing by adjusting their routines accordingly. They develop strategies to generate a

sense of mutuality, social presence and inclusion: for instance, they address different sites individually, and collect feedback in rota from each site. For instance, a teacher who used to move a lot did not wander as far from the lectern as he used to do in the collocated settings. Therefore, he did not slip out of the remote student's view. Yet, his highly informational gesticulation remained, and his temporal articulation was ideally paced for video-mediated lecturing; for instance, he slightly prolonged a pause when a technical mode switch was taking place.

The settings were not furnished with particular pointing functionality. Therefore, they were not capable of facilitating effective across-site gestural deixis (as did the experimental settings in Case study 1). Had there been several hands up at the remote site simultaneously, the teacher' hand gesture towards the remote audience would not have been effective as a floor control function.

Yet, the settings facilitated to some extent *procedural cuing* that supports *floor control* and *turn taking*: if any of the students wishes to signal by raising the hand that s/he has got a question, the teacher can easily spot it with a sweeping glance at the audience. Similarly, if a parallel frame is displayed of the teacher in his/her situational context, a remote student is able to capture a wider range of cues than a mere cursor movement on the focus of communication.

The settings implemented in Case study 2 provided a reasonably natural-resembling *reciprocity of perspectives* with a multi-site audience[115]. They reduced inequality between the local and the remote students in reference to the teacher's visual attention (motivational impact of *mutuality*), supporting thereby a sense of *social presence* at the remote site. They enabled teacher/student interaction across locations, and by visual procedural cues, they supported across-site interaction and turn taking. Yet, a member of AV staff is required to operate the controls, in particular during the Q&A sessions. As no pointing functionality was made available, it was not possible to point in an effective way, with the aim to guide remote participants' attention, to objects that were physically available at the *remote* site.

6.6 Design Principles for Hybrid Interaction

Based on the findings from Case study 2, it is concluded that the deixis-based design principles (See also Table 15) successfully guided the orchestration of the complementary

[115] The fact that there occurred technical disturbances do not undermine the spatial design principles.

role perspectives and provided functional affordances for video-mediated lecturing. In terms of spatial design, the hybrid settings provided

- *the scope* of shared information space (interaction space) that covers
- a range of *topic*-relevant objects so that
- the attention of the audience can be guided to an object of reference (*focus* in the interaction space)
- in a task-relevant way (*mode*).

The exact *focus* of pointing is, of course, in the teacher's situational control (See also Figure 10, Table 4).

6.7 Summary

The findings from the two case studies that were conducted to validate the framework in spatial design were reported. In both cases, a technical connection was considered an artificial extension of the person's information space. The concept of deixis (Figure 9), the human-environment interface and the human-human interface were applied in the articulation and interrelation of the participant perspectives at different levels of coordination (Figure 10).

Case study 1 helped to describe the mechanism of multimodal deixis in terms relevant to workplace design. The findings from the usability tests of the experimental test settings suggested that the interaction space conveyed relevant information, in sufficient amount, in order for the test users to establish shared understanding of deictic reference (=infer the connection between the deictic centre (pointing hand) and the object indicated). In other words: the settings enabled the communicative function of an across-site pointing gesture. The test users also flexibly adapted to an unfamiliar frame of orientation and interpretation.

The findings from Case study 1 (user experience of the settings for video-mediated lecturing), suggest viability of the design principles derived from the framework (See Chapter 4): based on reciprocity of perspectives and visual common ground, and orchestrated according to the cognitive ergonomic requirements, they helped to design functional affordances for effective interaction and communication across distributed sites (effective deixis, social presence, effortless adoption of functionalities, and a flexible adaptation to the frame of orientation and interpretation).

In a practical case (Case study 2), the design principles were informing the layout of the hybrid settings for video-lecturing: the teacher's and the student's role perspectives were spatially coordinated to provide visual[116] mutuality with local and remote participants. A technology-mediated connection was regarded as an audio-visual extension of the collocated settings.

In the design of functional affordances for distributed communication and collaboration, the main concerns are

- information relevant to the common purposes, and
- communication mode available to / required for effective joint efforts of the participants.

The concepts applied to designing the experimental test settings in Case study 1 helped in Case study 2 to define features for across-site communication in terms of

- informative-ness (topic),
- attention management (focus),
- across-site mutuality (social presence), and
- turn-taking (floor control)
- reciprocity of perspectives (speaker – addressee).

For the lecturer, the implemented interaction space enabled visual feedback from the overall audience. While disambiguation of the referent is frequent in communicating a complex topic, the layout made it easy for the teacher to guide the whole audience's attention to the object of communication. The interaction space also enhanced also a feeling of inclusion and social presence across distributed classes.

In all, Chapter 5 shows that a deixis-based, affordance-oriented design approach is applicable to hybrid type of work settings (=setting where people communicate simultaneously with collocated participants in co-presence and with remote-site participants in video-mediated presence). The outcome of Chapter 5 also points towards a shift in workplace design paradigm: the point of departure in hybrid interaction design is in the human-human interface (HHI). In the coordination of participant perspectives, the concept of deixis is a helpful investigation and coordination tool.

[116] Audial mutuality was asymmetric as the microphone was muted at the remote end during the lecturer's presentation phase, whereas during Q& A session, it was turned on.

7 Conclusions

This chapter summarises and concludes the thesis by discussing its contribution in terms of originality, significance and limitations. In addition, it outlines further research topics related to this research.

The thesis builds on multidisciplinary research in communication and collaboration (Chapter 2, Chapter 3). It shows how understanding of the mechanism of deixis is important for the design of 'hybrid' work settings (=settings where Information and Communication Technologies are an integral part) (Chapter 4). It is motivated by the context change that the penetration of mobile and Internet technologies has brought about over the past few decades to human practices.

Space appears in multidisciplinary research literature as the foundational grid for things to take place, forms get shape, phenomena to be situated, and for impressions / appearances to be made. In the thesis, space is considered basic (and pervasive) infrastructure that provides common ground for people and objects, informational resource for human cognition, and a medium for information flow between different perspectives (mutuality and reciprocity of perspectives)

- in natural communication, and
- through ICT-mediation as artificial extension of natural interaction space.

This thesis takes communicative situation as a unit of analysis. Goffman (op.cit.) in particular pointed out the importance of the 'neglected' situation in communication studies. Buehler's (1934) I-here-now again conceptualizes the origo of subjective orientation. Tversky's (op.cit.) research on spatial cognition and spatial navigation of humans helped to describe a person's perspective onto the world on one level as physical/spatial, on another one, as cognitive. Clark's (op.cit.) contribution theory of conversation and common grounding at the language level provided conceptual tools for the coordination of individual perspectives on a communicative situation. Interpersonal communication is a social encounter where two or more such perspectives are interconnected; in collaboration, the participants' contributions are coordinated through turn taking. Roth's (op.cit.) studies of communication and collaboration helped to grasp the role of inter-modality and to take it into consideration when developing the framework for spatial design. Gibson's concept of affordance from Perceptual Psychology helped to spell out the interface of a communicating person with the environment. In connection with Clark's concepts, the interface could be categorized into

two different types, Human-environment Interface and Human-Human Interface. Roth's (2004) inter-modality as a concept helped to cover also technology-enhanced interface. Diessel's (op.cit.) studies of demonstratives and declarative pointing in human communication helped to understand the role of pointing gesture also from the spatial design point of view.

Multimodal deictic practices are part of human communication and collaboration. From spatial design perspective, multimodal deixis requires powerful investigation tools. The thesis proposes a deixis-based approach and framework for workplace design that integrates perspectives of Architecture (spatial layout of components), Cognitive Psychology (cognitive infrastructure and patterns of social communication as constraining frames), and Human-Computer Studies (configuration of functional affordances to extend natural interaction space). The framework takes into account spatial behaviour, intentional action, and modal resources of a communicative situation.

The interaction space of communicating participants is composed of different levels of phenomena, each governed by particular laws/regularities (Figure 4). Therefore, three relational concepts are employed to describe structural / functional / communicative / cognitive aspects of communicative situation in order to inform design. They are

- human-environment interface (HEI), and
- human-human interface (HHI)
    - use of tools as extension of human capacities: the interface of human with tool, here in particular human-computer interface (HCI).

Applying the three constructs, the thesis identifies four organisational nodes that are relevant to the coordination of the participants' perspectives and to the coordination of their contributions; they are locus, modus, topic and focus.

The capacity of the interaction space to facilitate shared understanding is described as a function of

- the spatial position/cognitive state of the participants,
- connection (in functional/sensory terms), and
    - technology as a component of the connection.

7.1 Originality

As a whole, the thesis positions itself between Architectural Design and Interaction Design. In such consideration, it contributes to place, interface, affordance and interdependence discourse, and to design methods in the area of hybrid interaction and cross-modality.

The thesis explains gestural deixis for workplace design. Thereby, it provides an enriched structure of the context of interpersonal meaning making.

It presents an original approach that brings together three different approaches, that is, Architecture, Cognitive Psychology and Human-Computer Interaction, to investigate human navigation, communication and collaboration with the aim to inform spatial design.

Its main contribution is in introducing *situated deixis* as an investigation and design tool in spatial design related research literature.

Methodologically, the thesis combines cognitive and ethnographic intellectual traditions in a complementary way into a method to analyse the flow of communication as a spatial performance.

A methodological contribution is in bringing demonstratives and locatives into usability studies and introducing them as analytic tools in the research of communication and collaboration that aims to inform spatial design.

The particular strength of deixis is that, when situated, it encompasses the dimensions of human navigation, communication and collaboration that are relevant to spatial design (settings for communication and collaboration). Situated deixis also encompasses modal aspects of communication. That is a crucial bonus because modal fragmentation complicates the design of technology enhanced affordances for collaboration. Thereby, the thesis points to a paradigm shift in spatial workplace design.

Taking a communicative situation as a unit of analysis, the thesis considers technology-mediated connection an artificial extension of natural human information space. In the deixis-based design approach, two concepts, the human-environment interface and the human-human interface are applied with the concept of deixis. Complementary role perspectives can be orchestrated to provide an interaction space that enables topic related communication and collaboration in task relevant modes.

In this thesis, the term ‘framework’ appears somewhat different from what is the case in Computer Science: it is used here rather as a prism to describe phenomenal complexity of deictic practices.

The framework is based on a scientific explanation of deixis that takes into account both physical to cognitive phenomena, but it is additionally a description of deixis that *renders the design-relevant features of human-human interface in order to guide the configuration of work setting*s. It thus helps to articulate and organise communicative requirements for workplace design and translate them into spatial-functional terms. It is in such terms that design operates when composing/configuring functional affordances for communication and collaboration. In this thesis, mutuality is viewed as an operational connection that ranges from the physical level to that of symbolic communication and intentional action between communicating participants in a communicative situation (Table 15, Figure 4, Figure 5).

Thereby, deixis-based design approach points out a shift in workplace design paradigm motivated by the worldwide adoption of communications technologies and their impact on human practices.

The practical design of technology-enhanced workplace involves configuration of functional affordances for such a connection between communicating perspectives (=participants) that is coordinated and customised at four different levels, each foundational for human communication:

- Locus of participants (where?)
- Modus of co-operation (how?, that is, the set of modalities/channels involved/required in co-operation)
- Topic of co-operation (what? as shared frame of reference)
- Focus of co-operation (what? as co-attentional focus of co-operation).

Though this thesis only focuses on *gestural* deixis, the concept of deixis covers human sensory modalities and motor (operational) actions.

As an outcome, the thesis contributes to workplace design-related research: it provides

- a framework that incorporates people, place and process perspectives on collaboration to investigate deixis,
- it identifies deictic gesture as key object in the research in deixis and cross-modality,

- it situates reciprocity of perspectives in communicative situation in order to describe the context of communication in a relevant way for workplace design,
- it combines cognitive and (video)ethnographic approaches to explain communicative function of deictic gesture for workplace design, and
- it outlines a deixis-based design approach that guides effective coordination of the two foundational relations of collaboration: human-environment interface (HEI) and human-human interface (HHI).

7.2 Relevance

The approach described in this thesis is in response to the recognised need to develop approaches that will cover requirements of interaction in the hybrid settings, (ICT-enhanced interaction space). The practical motivation arises from the pace of the technological context change: the diffusion of the novel communication tools has been particularly fast since the turn of the 21st century[117] [118].

The prior inventions and their diffusion - such as the adoption of writing, printing technology, modern transport systems, and electronic mass media - give some idea of the scale of the implications that technological change has on individuals and communities. The consequences unfolding from the recent context change on humanity and on the environment will unavoidably be more overwhelming.

The extent of the context change challenges also workplace design to move from a traditional design paradigm towards a more integrated paradigm in order to cover the diversity and fluidity in the workscape in the 21st century: the context of interpersonal meaning making in the hybrid settings requires a formulation that covers both co-presence and virtual presence *within a single framework*.

Discipline-specific approaches to a communicative situation cover complementary aspects relevant to the empirical research of deictic practices, and to the spatial design of the settings. Yet, none of them is sufficient alone (due to the limited scope of validity) to explain the context of interpersonal meaning making in the hybrid settings. However, if bridged in a consistent way, the compilation of such arguments and complementary approaches provides a wider scope of validity than any one of them alone: a

[117] http://www.internetworldstats.com/top25.htm http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf
[118] http://www.bbc.co.uk/news/10569081 (Accessed July 26, 2010)

multidisciplinary explanation of deixis accommodates both *reciprocity of perspectives* and *modal integration* [119] within a single framework.

By combining traditional workplace design with interaction design, the thesis outlines a design paradigm (hybrid interaction design'): it brings together Architecture, studies of Human Communication and Human-Computer Interaction developing a deixis-based, affordance-focused approach that situates the human perspective *in a communicative situation*.

- Thereby, it shifts the focus from an anthropocentric perspective (that is, 'the Vitruvian man') to *mutuality* (>reciprocity of individual perspectives);
- it moves from a traditional design paradigm to an integrated paradigm, presenting design principles that are applicable to any type of settings.

7.3 Limitations

While the practical motivation of this research was workplace design, the thesis focused on spatial constraints of collaboration, in particular on deictic practices as the key object. Therefore, the strength of this thesis is on a theoretical side, in bringing together a number of theories that help understand deixis in practical design. The description of deixis is heavily anchored in existing theory, showing that a multidisciplinary lens gives a view on deixis in such a way that we can build hypothetical assumptions for exploration and test them through experimental designs. From the theoretical point of view, such exploration is backed with several partial and complementary theories.

The weaker side of the thesis is empirical. From evidence point of view, our video-data can be considered hard data in the sense that particular (spatial) relations can be captured in measurable terms. Video-recorded data shows undeniably that the assumption of the underpinnings of deictic reference in video-mediated settings held ground in our usability tests; yet, further and more focused research is needed with a sufficient number of users, using also statistical backing. Findings from video-data have in this research been supported by narrative evidence (such as confirmation of the users in feedback discussion and accordance of video-data with users' written comments on a short questionnaire).

One of the empirical limitations of the thesis is also that the framework was evaluated only in audio-visual-gestural domain. For instance, remote manipulation of a tacit object would require haptic modality, which was out of the scope of this project and would

[119] Reciprocity of perspectives implies two spatial origos, and modal integration implies such coordination of individual perspectives that enables a referential triangle.

require additional functionality in test bed application. Yet, reciprocity of perspectives as explained in Chapter 4 (Figure 10) is supported in multidisciplinary theory: it is in accordance with Figure/Ground law in Gestalt theory (Koffka, 1935), Frame theory (Goffman (1974), situated cognition (e.g. Roth, 2004), situated social cognition (e.g. Gabora, 2008), and cognitive linguistic view on common ground (Clark, 1996). While considering the overall interface of an actor with the environment as affordance, it has connection to ecological views on affordance (Gibson, 1979; Hartson, 2003; Turner, 2005). On such basis, the framework has potential to accommodate any modality of human communication.

There are several specific limitations of the first case study. First, the user assumption of the first case study was someone involved in the working life, that is, an adult person who has normal sensory-motor faculties. Therefore, the point of view in studying deictic practices is biased because constraints resulting from disabilities, aging and temporary impairment of sensory-motor and cognitive faculties were not taken into account.

Second, the technical delay in signal transmission and its impact on inferring deictic reference in the technology-mediated settings was out of scope of the project. In the test setting, the participants had audio co-presence while working in a video-mediated way. In addition, lighting conditions were not considered in any depth, neither was the role of microphone positions considered in the spatial layout. Furthermore, only one pointing device was used in the first round of the tests, and one pointing device at both ends in the second round. Therefore, it was not possible to consider situations where more than two people were pointing at the same time.

Yet, we can undeniably infer from our video-data that manipulation of attention with mediated pointing supports the identification of the object of reference (and thereby, shared understanding) in video-mediated communication. We have practical evidence to justify further research based on our theoretical assumption of spatial-functional-social-cognitive constitution of mutual understanding.

The specific limitations of the second case study are the following: The practical design case focused on implementing such a teacher-student interface that would accommodate local and remote communication at the same time. As this thesis focuses on the methodological side of the project, it only reports the teacher's perspective, omitting students as another key stakeholder perspective. Yet, the method of approaching student's point of view is, in principle, similar to the one applied to the teachers: observation,

confirmation of observations in feedback discussion, focus group discussions and simulations.

From the design point of view, also other stakeholder groups equal to teachers should be covered: it became clear in the course of the project that one of the critical links is *horizontal* communication between different stakeholder groups, in order to provide an efficient learning interface. It is in that respect imperative for teachers, students and AV staff to learn and benefit from mutual learning process, developing best practices in the video-mediated learning settings; moreover, refining them in the course of practice, and informing further design and development of learning interfaces. In such respect, the limitation of this thesis is that the data of emerging practice is minimal. Yet, the methodological aspect has been covered.

Another limitation of the second case study was the type of activity: it focused only on learning environment. Furthermore, the report in this thesis only covers one of the distributed locations, leaving the remote ends out of scope. The justification of this is that the remote sides remained in the state they were in the beginning of the project, whereas the reported location was rearranged in the reported manner.

7.4 Further Research

The thesis points to several further research topics. Among them is a more thorough study of cross-modality in video-mediated communication, approached from a multidisciplinary angle. For instance the role of focus and direction of gaze, as well as intonation of voice, are feasible topic areas to focus on in a research that seeks to clarify expressions of motivation and their recognition for design purposes.

Another topic for multidisciplinary research is navigation in Mixed Reality. In order to better understand spatial reasoning and adaptation when people move from a familiar context to an unfamiliar one, for instance from plain videoconferencing to Mixed Reality videoconferencing, contribution from neuropsychological and neurocognitive research and related methods would be required.

As this research enabled only one pointer per site, a further experimental design topics include particular scenarios, such as multi-user pointing, that is, simultaneous use of pointing functionality by multiple participants in videoconferencing; a case could be for instance voting between optional locations on a map. In a more detailed case, the research question would focus on pointer-reference vector (in terms of identification of who is pointing what), when several pointing representations appear at the same time.

From the technical point of view, sufficiency/excess condition of mediated pointing to convey communicative function is an interesting question. Another technical topic is relevance pattern recognition: for instance, how to provide camera view on objects of relevance during a lecture. A far more important topic than the above mentioned questions is the accelerating context change in consideration with human wellbeing.

Another scenario for future research is floor control in a video-mediated conference where the chair is supported by additional information tagged to the remote participants.

Further research questions include for instance mediated pointing in remote guidance in specific fields of practice (e.g. health care, industry), differences in collocated and video-mediated pointing behaviour, and etiquette in video-mediated pointing. Taking the user perspective, user preferences provide additional topics for research in terms of pointing device, its usability, as well as pointing techniques applied by different users.

This research was conducted at the time when we probably do not understand the vast implications that the emergence of Internet and mobile technologies is bringing about: only ten years ago, there were not billions of people using the Internet and mobile phones as there are at the time of writing this thesis. In such a context, a technology critical attitude is not only necessary in design (and design education in particular) but also in a wider societal consideration; in the world around us, whole economies, societies, and communities struggle to cope with a rapid context change, and individual people suffer from multiple stress-related symptoms. An adaptation to a context change requires active *relating to* and *reflection of* the situation; that is, *distancing* in order to get *a bigger picture* of the situation, to keep *in a sound balance* and *on the right track* (not *getting lost*). In other words: the faster the context changes, the more pressure people (and whole communities) face in their situational (human-environment) interface[120].

The context change implies a specific problem that arises from two different types of process: the pace of the biological/physiological processes cannot be accelerated beyond their internal tolerance whereas the pace of cultural processes keeps accelerating along with increasing flows of goods, people, information and innovations. Though human coping mechanisms are flexible, they have their limits, too, as do the globe's natural resources. Therefore, *the technological change as it has accumulated over time in the present-day human-environment interface* opens up deep-going questions concerning

[120] It implies even wider/deeper-going questions of interdependence and sustainability, existence, bounded-ness, and difference: an adequate response to the context change is the condition for humanity to navigate in its ecology and survive. Learning from prior experiences provides an internal compass for navigation.

consumerism, environmental responsibility, and ‘responsible innovation’. Such topic needs a place in design education, and it should have place in the research agenda of a multidisciplinary research community.

Bibliography

Adams, J., Rogers, B., Hayne, S., Mark, G., Nash, J. & Leifer, L. 2005. 'The Effect of a Telepointer on Student Performance and Preference'. *Computers and Education*, 44 (1), 35-51.

Akman, V. 2000. 'Rethinking context as a social construct'. *Journal of Pragmatics*, 32, 743-759.

Alexander, C. 1977. *A Pattern Language: Towns, Buildings, Construction.* USA: Oxford University Press.

Aristotle. 2007. *Physics,* Book 4. Translated by Hardie, R. P & Gaye, R. K. eBooks@Adelaide. Available at http://ebooks.adelaide.edu.au/a/aristotle/physics/book4.html (Accessed 21.09.2011).

Austin, J. L. 1962. *How to Do Things with Words*. Oxford University Press.

Barakonyi, I., Fahmy, T. & Schmalstieg, D. 2004. 'Remote collaboration using augmented reality videoconferencing. *Proceedings of Graphics Interface 2004*, 89–96. (London, Ontario, Canada, May 17–19, 2004). ACM International Conference Proceeding Series, vol. 62. Canadian Human–Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario.

Benevolo, L. 1977. *History of Modern Architecture*. MIT Press.

Billinghurst, M. & Kato, H. 2002. 'Collaborative Augmented Reality'. *Communications of the ACM*, 45 (7), 64-70.

Bloomer, K.C. & Moore, C.W. 1977. *Body, Memory, and Architecture.* New Haven and London: Yale University Press.

Borchers, J. O. 2001. 'A pattern Approach to Interaction Design'. *AI & Society*, 15, 359-376.

Bratman, M. E. 1992. 'Shared Cooperative Activity'. *The Philosophical Review*, 101 (2), 327-341.

Bühler, K. 1934/1990. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.

Burke, K. 1945. *A Grammar of Motives*. New York: Prentice-Hall.

Ching, F. D. K. 1979. *Architecture: Form, Space, and Order.* New York: Van Nostrand Reinhold.

Ciolfi, L. 2004. *Situating 'Place' in Interaction Design: Enhancing the User Experience in Interactive Environments.* PhD Thesis, University of Limerick.

Clark, H. H. 1996. *Using language.* Cambridge: Cambridge University Press.

Clark, H. H. 2001. ‘Conversation’. Smelser, H. J. & Baltes, N. J. eds. *Inter-national Encyclopaedia of Social and Behavioral Sciences*, 2744-2747. London: Elsevier.

Clark, H. H. 2005. ’Coordinating with each other in a material world’. *Discourse and Society*, 7(4-5), 507-525.

Clark, H. H. & Brennan, S. A. 1991. ‘Grounding in communication’. Resnick, L.B., Levine, J. M. & Teasley S. D. eds. *Perspectives on socially shared cognition*. Washington: APA Books.

Clark, H. H. & Wilkes-Gibbs, D. 1986. ‘Referring as a collaborative process’. *Cognition,* 22, 1-39.

Clark, H. H. & Schaefer, E. F. 1989. ‘Contributing to Discourse’. *Cognitive Science,* 13, 259-294.

Clark, H. H. & Krych, M. A. 2004. ‘Speaking while monitoring addressees for understanding’. *Journal of Memory and Language,* 50, 62-81.

Clark, H. H., Schreuder, R. & Buttrick, S. 1983. ‘Common Ground and the Understanding of Demonstrative Reference’. *Journal of Verbal Learning and Verbal Behaviour,* 22, 245-258.

Daft, R. L. & Lengel, R. H. 1986. ‘Organizational information requirements, media richness and structural design’. *Management Science*, 32 (5), 554-571.

Danesi, M. ed. 2000. *Encyclopedic Dictionary of Semiotics, Media, and Communications*. Toronto: University of Toronto Press.

Darwin, C. 1872. *The expression of the emotions in man and animals*. London: John Murray.

Détienne, F. 2006. ‘Collaborative design: Managing task interdependencies and multiple perspectives’. *Interacting with Computers*, 18, 1-20.

Dewey, J. 1924. *Democracy and Education*. New York: Macmillan Company.

Diessel, H. 2006. ‘Demonstratives, joint attention, and the emergence of grammar’. *Cognitive Linguistics,* 17 (4), 463–489.

Diessel, H. 2003. ‘The relationship between demonstratives and interrogatives’. *Studies in Language*, 27 (3), 635-655.

Dijk, A. van, 2006. ‘Discourse, context and cognition’. *Discourse Studies,* 8(1), 159-177.

Dix, A., Finlay, J., Abowd, G. D. & Beale, R. 2004. *Human Computer Interaction.* Third edition. Prentice Hall.

Dourish, P. & Bellotti, V. 1992. ’Awareness and Coordination in Shared Work Spaces’. Proceedings ACM Conference on *Computer-Supported Cooperative Work CSCW’92.* Technical Report EPC-92-101, 1-9.

Dourish, P. 1997. ‘Extending Awareness Beyond Synchronous Collaboration’. Position paper for *CHI’97 Workshop on Awareness in Collaborative Systems.* Atanta, Georgia, March 1997.

Dourish, P. 2006. ‘Re-Spacing Place: Place and Space Ten Years On’. Proc. ACM Conf. *Computer-Supported Co-operative Work CSCW 2006* Banff, Alberta, 299-308.

Eccles, J. C. & Popper, K. 1984. *The Self and Its Brain: An Argument for Interactionism.* Routledge.

Elias, N. 1982. *Über die Einsamkeit der Sterbenden in unseren Tagen*. Frankfurt am Main: Suhrkamp. (Published in English as *The Loneliness of the Dying*, Oxford: Blackwell, 1985).

Elias, N. 1991. *The Symbol Theory.* Sage Publications.

Enfield, N. J., Kita, S. & de Ruiter, J. P. 2007. ‘Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, 39 (10), 1722-1741.

Flender, C. (date not available) ‘Dynamic Co-Emergence as a Way of Being-in-the-World’. Available at http://sky.fit.qut.edu.au/~flender/coemergence.pdf (Accessed 29.09.2011).

Flender, C., Kitto, K. & Bruza, P. 2009. ‘Nonseparability of Shared Intentionality’. *Quantum Interaction.* Lecture notes in Computer Science vol. 5494/2009.

Fischer-Lichte, E. & Wulf, C. eds. 2008. *Proceedings of the Berlin Ritual-Conference*, Dec. 5. 2008.

Fitzpatrick, G., Kaplan, S., and Mansfield, T. 1996. ‘Physical spaces, virtual places and social worlds: A study of work in the virtual’. Proceedings of *CSCW'96*, 334-343. ACM Press, Boston, MA.

Flinn, M., Gearny, D. C. & Ward, C. V. 2005. ‘Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence’. *Evolution and Human Behavior*, 26, 10-26.

Focault, M, 1977. *Discipline and Punish: the birth of the prison.* New York: Random House. [*Surveiller et Punir: Naissance de la prison* (1975) Paris: Gallimard.]

Freksa, C. et al. eds. 2003. *Spatial Cognition III*, LNAI 2685. Berlin Heidelberg: Springer-Verlag.

Freud. S. 1972. *Ahdistava kulttuurimme*. [*Das Unbehagen in der Kultur*, 1930.] Translation into Finnish by Erkki Puranen. Helsinki: Weilin + Göös.

Fullwood, C. & Doherty-Sneddon, G. 2006. ‘Effect of gazing at the camera during a video link on recall’. *Applied Ergonomics*, 37, 167–175.

Gabora, L. 2008. 'The cultural evolution of socially situated cognition'. *Cognitive Systems Research*, 9 (1-2), 104-113.

Gaver, W. 1991. 'Technology affordances'. In Proceedings of the *CHI 1991*, 79 – 84. ACM Press: New York.

Gaver, W. 1992. 'The Affordances of Media Spaces for Collaboration'. Proceedings of *CSCW '92*, 17-24. ACM Press.

Gaver, W. 1996. 'Affordances for interaction: The social is material for design'. *Ecological Psychology*, 8 (2), 111-129.

Geertz, C. 1973. *The Interpretation of Cultures.* New York: Basic Books.

Gemmell, J., Toyama, K., Zitnick, C. L., Kang, T. & Seitz, S. 2000. 'Gaze Awareness for Video-Conferencing: A Software Approach'. IEEE MultiMedia, 7 (4), 26-35.

Gergle, D., Kraut, R. E. & Fussell. S. R. 2004. 'Action as Language in a Shared Visual Space'. *CSCW'04,* November 6-10, 2004, Chicago, Illinois, USA.

Gergle, D., Rose, C. P., Kraut, R. E., 2007. 'Modeling the impact of shared visual information on collaborative reference'. Proceedings of *CHI 2007,* 1543–1552. ACM Press, New York.

Gibson, J.J. 1977. 'The Theory of Affordances'. Shaw & R. J. Bransford, eds. *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, 67-82. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Gill, S. P. (Date not available). 'Engagement Space: The Body in Parallel and Sequential Time and Space'. Available at http://www.utexas.edu/coc/cms/International_House_of_Gestures/Conferences/Proceedings/Contributions/Gill/Gill.html (Accessed 15.03.2006).

Gill, S. P. & Borchers, J. 2004. 'Knowledge in Co-Action: Social Intelligence in Collaborative Design Activity'. *AI & Society,* 17 (3-4), 322-339.

Goffman, E. 1956/1959. *The Presentation of Self in Everyday Life*. University of Edinburgh Social Sciences Research Centre.

Goffman, E. 1964. 'The neglected situation'. *American Anthropologist, New Series,* 66 (6) *Part 2: The Ethnography of Communication* (Dec. 1964), 133-136. Blackwell Publishing.

Goffman, E. 1974. *Frame analysis: An essay on the organization of experience.* London: Harper and Row.

Goodwin, C. 2000. 'Action and embodiment within situated human interaction'. *Journal of Pragmatics,* 32, 1489-1522.

Gutwin, C. & Greenberg, S. 2004. 'The importance of awareness for team cognition in distributed collaboration'. Salas, E. & Flore, S. M. eds. *Team cognition: Understanding the factors that drive process and performance*, 177-201. Washington: APA Press.

Hall, E. T. 1966 /1990. *The Hidden Dimension*. Reissue edition. Anchor.

Harrison, A., Wheeler, P. & Whitehead, C. eds. 2003. *The Distributed Workplace: Sustainable Work Environments.* Routledge.

Harrison, S. & Dourish, P. 1996. 'Re-Place-ing Space: The Roles of Place and Space in Collaborative Systems'. *Proceedings of 1996 ACM Conference of Computer supported cooperative work*, 67-76.

Hartson, H. R. 2003. 'Cognitive, physical, sensory, and functional affordances in interaction design'. *Behavior and Information technology,* 22 (5), 315-338.

Harvey, D. 2004. 'Space as a key word'. Paper for *Marx and Philosophy Conference*, 29 May 2004, Institute of Education, London.

Heath, C. & Hindmarsh, J. 2002. 'Analysing Interaction: Video, Ethnography and Situated Conduct'. May, T. ed. *Qualitative Research in Action*, pp. 99-121. London: Sage Publications.

Heath, C. & Luff, P. 1991. 'Disembodied Conduct: Communication through video in a multi-media office environment'. *CHI'91. Proceedings of the SIGCHI Int. Conference on Human Factors in Computer Systems: Research through Technology.* ACM.

Heath, C., Luff, P., Kuzuoka, H. & Yamazaki, K. 2001. 'Creating Coherent Environments for Collaboration'. Proceedings of *ECSCW 2001*, 119-128. Kluwer Academic Publishers.

Heidegger, M. 1971. 'Building, Dwelling, Thinking'. Hofstadter, A. *Poetry, Language, Thought*, translated by Albert Hofstadter, Harper Colophon Books, New York. Available at http://students.pratt.edu/~arch543p/readings/Heidegger.html (Accessed 28.05.2011)

Hillier, B. & Hanson, J. 1984. *The Social Logic of Space.* Cambridge: Cambridge University Press. Reprint edition 1989.

Hindmarsh, J. & Heath, C. 2000. 'Embodied reference: A study of deixis in workplace interaction'. *Journal of Pragmatism,* 32 (12), 1855-1878.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. & Greenhalgh, C. 1998. 'Fragmented interaction: Establishing mutual orientation in virtual environments'. ACM conference on *Computer-Supported Cooperative Work (CSCW'98)*, 217-226. ACM Press, November 1998.

Hinds, P. J. & Kiesler, S. eds. 2002. *Distributed work*. USA: The MIT Press.

Hofstadter, A. 1971. *Poetry, Language, Thought.* Translated by Albert Hofstadter. New York: Harper Colophon Books.

Holyoak K. & Morrison, R. eds. 2005. *Handbook of Reasoning*. Cambridge: Cambridge University Press.

Hutchins, E. 1995. *Cognition in the Wild.* MIT Press.

Iacoboni M., Molnar-Szakacs I., Gallese V., Buccino G., Mazziotta J. C. & al. 2005. 'Grasping the Intentions of Others with One's Own Mirror Neuron System'. *PLoS Biol* 3 (3): e79.

Iedema, R., Long, D., Carroll, K., Stenglin, M. & Braithwaite, J. 2006. 'Corridor work: how 'liminal' space can be a focal resource for handling complexities of multi-disciplinary health care'. Muetzelfeld, M. ed. *Proceedings of the 11th International Colloquium of the Asia-Pacific Researchers in Organization Studies (APROS)*, 238-247. 4-6 December 2005, Melbourne.

Ilyenkov, E. 1977. *Problems of Dialectical Materialism.* (Translated by A. Bluden). Progress Publishers.

International Telecommunication Union. 2010. Measuring the Information Society 2010. Available at http://www.itu.int/ITU-D/ict/publications/idi/2010/Material/MIS_2010_without%20annex%204-e.pdf (Accessed 21.07. 2010)

Internet World Stats. 2010. Available at http://www.internetworldstats.com/stats.htm (Accessed 06.07.2011) http://www.internetworldstats.com/top25.htm (Accessed 04.03.2011)

Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S., Matsushita, Y., 1995. 'MAJIC videoconferencing system: experiments, evaluation and improvement'. Marmolin, H., Sundblad, Y., Schmidt, K. eds., Proceedings of the Fourth Conference on *European Conference on Computer-Supported Cooperative Work*, 279–292. (Stockholm, Sweden, September 10–14, 1995). ECSCWKluwer Academic Publishers, Norwell, MA.

ITU 2011. *ICT Facts and Figures: The World in 2011.* Available at

http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf (Accessed 15.06.2012)

Isaacs, E. A. & Tang, J. C. 1994. 'What Video Can and Cannot Do for Collaboration: A Case Study'. *Multimedia Systems*, 2, 63-73.

Jaimes, A. & Sebe, N. 2003. 'Multimodal human computer interaction: a survey'. Proceedings of *IEEE International Workshop on HCI*, Genova, Italy, 2003.

Järvilehto, T. 2000. 'Consciousness as Co-operation'. *Advances in Mind-Body Medicine*, 16, 89-92.

Kendon, A. 1967. 'Some Functions of gaze-direction in social interaction'. *Acta Psychologica*, 26, 22-63.

Kendon, A. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters.* Cambridge University Press.

Knapp, M. I. & Daly, J. A. 2002. *Handbook of interpersonal communication*, 529-563. 3rd edition. Thousand Oaks, CA: Sage Publications.

Kock, N. 2005. 'Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools'. *IEEE Transactions on Professional Communication*, 48 (2), 117-130.

Koffka, K. 1935. *Principles of Gestalt Psychology*. New York: Hartcourt Brace.

Kolb. D. A. & Fry, R. 1975. 'Toward an applied theory of experiential learning'. C. Cooper ed. *Theories of Group Process.* London: John Wiley.

Kornberger, M. & Clegg, S. 2003. 'The Architecture of Complexity'. *Culture and Organisation*, 9 (2), 75-9.1

Koskinen, K.U., Pihlanto, P. & Vanharanta, H. 2003. 'Tacit knowledge acquisition and sharing in a project work context'. *International Journal of Project Management,* 21, 281–290.

Kostof, S. 1985. *A History of Architecture: Settings and Rituals.* New York. Oxford: Oxford University Press.

Kraut, R. E., Fussell, S. R., Brennan, S. E. & Siegel, J. 2002. 'Understanding the effects of proximity on collaboration'. Hinds, P. J. & Kiesler, S. eds. *Distributed work,* 137-162. USA: The MIT Press.

Kraut, R., Miller, M. and Siegel, J. 1996. 'Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication'. Proceedings of *CSCW '96*, 57-66. ACM Press.

Kravchenko, A. V. 2007. 'Essential properties of language, or, why language is not a code'. *Language Sciences,* 29 (5), 650-671.

Kuehn, G., Lievonen, M., Doerner, R. & Rosenberg, D., 2007. 'Augmenting the Information Space in Computer-Mediated Communication: Employing AR and Game Controller for Deictic References'. *Proceedings of the 10th International Conference on Humans and Computers 2007,* 30-44. HCC, Aizu, Japan.

Kuusisto, R. 2004. *Aspects of Availability: A teleological adventure of information in the lifeworld.* PhD Dissertation at Helsinki University of Technology.

Lefebvre, H. 1991. *Production of Space*. Blackwell Publishers.

Lievonen, M., Rosenberg, D., Dörner, R., Kühn, G. & Walkowski, S. 2009. 'Augmented Reality as Means for Creating Shared Understanding'. Norros, L., Koskinen, H., Salo, L & Savioja, P. eds. *ECCE 2009 - European Conference on Cognitive Ergonomics: Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments.* 161-168. Helsinki: VTT Symposium 258.

Lloyd McBurney, S. 2002. 'Pronominal reference in signed and spoken language: are grammatical categories modality-dependent?' Meier, R. P., Cormier, K. & Quinto-Pozos, D. eds. *Modality and Structure in Signed and Spoken Languages,* 329-269. Cambridge University Press, 2002.

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. 2003. 'Fractured Ecologies: Creating Environments for Collaboration'. Special Issue of the *HCI Journal: 'Talking About Things: Mediated Conversation about Objects'*, 18, 1 & 2 (2003), 51-84.

Luff, P., Hindmarsh, J. & Heath, C. eds. 2000. *Workplace Studies: Recovering work practice and informing systems design.* Cambridge: Cambridge University.

Lych, K. 1960. *The Image of the City.* MIT Press.

Malpas, J. 2002. 'The weave of meaning: holism and contextuality'. *Language and Communication,* 22, 403-419.

Mantovani, G. 1996. 'Social Context in HCI: A New Framework for Mental Models, Cooperation and Communication'. *Cognitive Science,* 20, 237-269.

March, E. G., Wales, R. & Pattison, P. 2006. 'The uses of nouns and deixis in discourse production of Alzheimer's disease'. *Journal of Neurolinguistics,* 19 (4), 311-340.

Marmolin, H., Sundblad, Y. & Schmidt, K. eds. 1995. Proceedings of the Fourth Conference on *European Conference on Computer-Supported Cooperative Work,* 279–292. (Stockholm, Sweden, September 10–14, 1995). ECSCW Kluwer Academic Publishers, Norwell, MA.

Markson, L. & Paterson, K. B. 2009. 'Effects of gaze aversion on visual-spatial imagination'. *British Journal of Psychology,* 100, 553-566.

Mauss, M. 1924/2000. *The gift.* W. W. Norton & Company. [*Essay sur le don*].

May, T. ed. 2002. *Qualitative Research in Action*. London: Sage Publications.

McGrenere, J. & Ho, W. 2000. 'Affordances: Clarifying and evolving a concept'. Proceedings of *Graphics Interface,* Montreal, May, 179-186.

McNeill, D. 2008. 'Gestures of Power and the Power of Gestures'. Fischer-Lichte, E. & Wulf, C. eds. *Proceedings of the Berlin Ritual-Conference*, Dec. 5. 2008.

McNeill, D. 2005. *Gesture and Thought.* Chicago: University of Chicago Press.

McQuiggan, S. W. & Lester, J. C. 2007. 'Modeling and evaluating empathy in embodied companion agents'. *International Journal of Human Computer Studies,* 65 (4), 348-360.

Mead, M. 1928. *Coming of age in Samoa.*

Meier, R. P., Cormier, K. & Quinto-Pozos, D. eds. 2002. *Modality and Structure in Signed and Spoken Languages.* Cambridge University Press.

Merleau-Ponty, M. 1945/1996. *Phenomenology of Perception.* Routledge. [*Phénoménologie de la perception* (1945). Paris: Gallimard].

Miller, G. A., Galanter, E. & Pribram, K. H. 1960. *Plans and the Structure of Behavior.* New York: Holt, Rinehart & Winston.

Morris, A. E. J. 1972. *History of Urban Form: Prehistory to the Renaissance*. New York: John Wiley & Sons.

Muetzelfeld, M. ed. 2006. *Proceedings of the 11th International Colloquium of the Asia-Pacific Researchers in Organization Studies (APROS)*. 4-6 December 2005, Melbourne.

Nardi, B. A. 2005. 'Beyond Bandwidth: Dimensions of Connection in Interpersonal Communication'. *Computer Supported Cooperative Work,* 14, 91-130.

Nielsen, J. 1993. *Usability Engineering.* San Francisco, CA: Morgan Kaufmann.

Nielsen, J. 1995. 'How to conduct heuristic evaluation?' Available at http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/

Niiniluoto, I. 1990. *Maailma, minä ja kulttuuri: Emergentin materialismin näkökulma*. [The title translated: 'World, I and Culture']. Helsinki: Otava.

Norberg-Schulz, C. 1980. *Genius Loci: Towards a Phenomenology of Architecture.* London: Academy Editions.

Norman, D. 1988/2002. *The Design of Everyday Things* (originally 1988: *The Psychology of Everyday Things*). Basic Books. 1st Basic Paperback.

Norman, D. A. 1999. 'Affordances, Conventions and Design'. *Interactions,* 6 (3), 38-43, May 1999. ACM Press.

Norros, L., Koskinen, H., Salo, L & Savioja, P. eds. 2009. *ECCE 2009 - European Conference on Cognitive Ergonomics: Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments.* Helsinki: VTT Symposium 258.

Nova, N. 2003. *Socio-cognitive functions of space in collaborative settings: a literature review about Space, Cognition and Collaboration.* CRAFT Research Report 1. Ecole Polytechnique Fédérale de Lausanne.

Ohno, T. 2005. 'Weak gaze awareness in video-mediated communication'. *CHI '05 Extended Abstracts on Human Factors in Computing Systems,* 1709–1712. (Portland, OR, USA, April 02–07, 2005). ACM, New York, NY.

Oksala, T. 1994. 'Ecological, Aesthetic and Ethical Unity in Architecture; Design: Ecology, Aesthetics, Ethics'. *Acta Polytecnica Scandinavica,* Ci 99, 23-27.

Pallasmaa, J. 1991. 'Culture, Identity, Environment'. *Rakennettu ympäristö - identiteetti – yhtenevä Eurooppa,* 107-118. Suomen Unesco-toimikunnan julkaisuja no 58. Helsinki: Yliopistopaino.

Pallasmaa, J. 2005. *The Eyes of the Skin: Architecture of the Senses.* John Wiley & Sons. 2nd Edition.

Peirce, C. S. 1868. 'On a New List of Categories'. *Proceedings of the American Academy of Arts and Sciences,* 7 (1868), 287-298. Available at http://www.peirce.org/writings/p32.html (Accessed 22.05.2011).

Piaget, J. 1950/1955. *The Child's Construction of Reality.* London: Routledge and Kegan Paul. [*La construction du réel chez l'enfant* (1950)].

Rasmussen, S. E. 1959. *Experiencing Architecture.* Cambridge, MA: MIT.

Rauhala, L. 1991. *Humanistinen psykologia.* Helsinki: Yliopistopaino.

Rauhala, L. 1993. *Eksistentiaalinen fenomenologia hermeneuttisen tieteen-filosofian menetelmänä*. Filosofisia tutkimuksia Tampereen yliopistosta, vol. 41.

Resnick, L. B., Levine, J. M. & Teasley S. D. eds. 1991. *Perspectives on socially shared cognition*. Washington: APA Books.

Robbins, P. & Aydede, M. eds. 2008. *Cambridge handbook of situated cognition.* Cambridge: Cambridge University Press.

Robertson, T. 2000. 'Building Bridges: Negotiating the Gap Between Work Practice and Technology Design'. *International Journal of Human-Computer Studies,* 53, 121-146.

Rogers, E. M. 1962. *Diffusion of Innovations*. Glencoe: Free Press.

Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. 1994. *Human Computer Interaction.* Addison Wesley.

Rosenberg, D. & Sillince, J. A. A. 2000. 'Verbal and nonverbal communication in computer mediated settings'. *International Journal of Artificial Intelligence in Education*, 11 (2000), 299-319.

Rosenberg, D., Foley, S., Kammas, S., & Lievonen, M. 2003. Interaction Space Theory: A Framework for Tool Development. *Proceedings of the International Symposium on Information and Communication Technologies (ISICT 2003)*, 446-451. Computer Science Press Trinity College Dublin, Ireland. Available at http://portal.acm.org/citation.cfm?id=963684

Rosenberg, D., Foley, S., Lievonen, M., Kammas, S. & Crisp, M-J. 2005. 'Interaction spaces in computer-mediated communication'. *AI and Society*, 19, 22-33. http://dx.doi.org/10.1007/s00146-004-0299-9

Rosenberg, D., Lievonen, M., Contreras, P., Murtagh, F., Kuehn, G. & Doerner, R. 2008. Application Design of Learning Grid in Computer-Mediated Communication. Salerno, S. et al. *The Learning Grid Handbook: Concepts, Technologies and Applications,* 107-123. IOS Press.

Roth, W-M. 2000. 'From gesture to scientific language'. *Journal of Pragmatics*, 32 (11), 1683-1714.

Roth, W-M. 2001. 'Situating Cognition'. *The Journal of the Learning Sciences*, 10 (1-2), 27-61.

Roth, W-E. 2002. 'From action to discourse: The bridging function of gestures'. *Cognitive Systems Research*, 3 (3), 535-554.

Roth, W-M. 2004. 'Perceptual gestalts in workplace communication'. *Journal of Pragmatics*, 36 (6), 1037-1069.

Salas, E. & Flore, S. M. eds. 2004. *Team cognition: Understanding the factors that drive process and performance*. Washington: APA Press.

SANE Project: Rosenberg. D., Crisp, M-J., Foley, S., Kammas, S. & Lievonen, M. (in public domain 2003). *SANE Human Environment Framework and Modelling.*

Saville-Troike, M. & Trudgill, P. 1982. *The Ethnography of Communication: An Introduction*. Blackwell Publishers.

Schober, M. F. 1993. 'Spatial perspective-taking in conversation.'. *Cognition*, 47 (1), 1-24.

Schutz, A. & Luckmann, T. 1973. *The Structures of the Life-World*. [*Strukturen der Lebenswelt*.] Translated by R. M. Zaner & H. Tristram Engelhardt, Jr. Evanston, IL: Northwestern University Press.

Schutz, A. 1973. *Collected Papers: The problem of Social Reality.* Vol. 1. Springer.

Shannon, C. 1948. 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 27, 379–423, 623–656.

Shaw R. & Bransford, J. eds. 1977. *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Short, J., Williams, E. & Christie, B. 1976. *The social psychology of tele-communications.* London, England: John Wiley.

Sillince, J. A. A. 1995. 'Shifts in focus and scope during argumentation'. *Journal of Pragmatics*, 25 (1995), 413-431.

Simonsen, J. & Kensing, F. 1997. 'Using Ethnography in Contextual Design'. *Communications of the ACM,* 40 (1997), no 7.

Smelser, H. J. & Baltes, N. J. eds. 2001. *International Encyclopaedia of Social and Behavioral Sciences*. London: Elsevier.

Smith, E. R. & Semin, G. R. 2007. 'Situated Social Cognition'. *Current Directions in Psychological Science,* 16 (3), 132-13.

Suchman, L. 1987. *Plans and Situated Actions: The Problem of Human-machine Communication.* New York: Cambridge University Press.

Talmy, L. 2003. 'Concept Structuring Systems in Language'. Tomasello, M. ed. *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure.* Volume 2, 15-46. Lawrence Erlbaum.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Harvard University Press.

Tomasello, M. ed. 2003. *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure.* Volume 2. Lawrence Erlbaum.

Tomasello, M. 2008. *Origins of Human Communication*. MIT Press.

Tomasello, M. 2009. *Why we cooperate.* Cambridge, Mass, London England: The MIT Press.

Tomasello, M. & Carpenter, M. 2007. 'Shared Intentionality'. *Developmental Science,* 10 (1) (2007), 121-125.

Torsti, M. 1988. 'Äitiys ja naisellinen itsetunto'. Hägglund, V. & al. eds. *Psykoanalyysin monta tasoa.* [The title translated: 'Multiple levels of Psychoanalysis']. Helsinki: Nuorisopsykoterapiasäätiön psykoanalyyttisia toimituksia II, 171-182.

Tuan, Y-F. 1977. *Space and Place: The Perspective of Experience*. University of Minnesota Press, Minneapolis.

Turner, P. 2005. 'Affordance as Context'. *Interacting with Computers,* 17 (6), 787-800.

Turner, P. 2008. 'Being-with: a study of familiarity'. *Interacting with Computers,* 20 (4-5), 447-454.

Tversky, B. & Martin Hard, B. 2009. 'Embodied and disembodied cognition: Spatial perspective-taking'. *Cognition,* 110, 124-129.

Tversky, B. 2003a. 'Navigating Mind and Body'. Freksa, C. et al. eds. *Spatial Cognition III,* LNAI 2685, 1–10. Berlin Heidelberg: Springer-Verlag.

Tversky, B. 2003b. 'Structures of Mental Spaces'. *Environment and Behavior,* 35 (1), 66-80.

Tversky, B. 2005. 'Visualspatial reasoning'. Holyoak K. & Morrison, R. eds. *Handbook of Reasoning,* 209-249. Cambridge: Cambridge University Press.

Tversky, B. 2008. 'Spatial cognition: Situated and embodied'. Robbins, P. & Aydede, M. eds. *Cambridge handbook of situated cognition.* Cambridge: Cambridge University Press.

Tversky, B., Bauer Morrison, J., Franklin, N. & Bryant, D. J. 1999. 'Three spaces of spatial cognition'. *Professional Geographer,* 5 (4), 516-524.

Vertegaal, R., Weevers, I., Sohn, C. & Cheung, C., 2003. 'GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 521–528. (Ft. Lauderdale, Florida, USA, April 05–10, 2003). CHI'03. ACM, New York, NY.

Vitruvius. Date not available. *The Ten Books of Architecture.* Available at http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Vitruvius/home.html http://www.hs-augsburg.de/~harsch/Chronologia/Lsante01/Vitruvius/vit_ar01.html (Accessed 31.03.2010) (Accessed 21.02.2011)

Walkowski, S., Doerner, R., Lievonen, M. & Rosenberg, D. 2011. 'Using Game Controller for Relaying Deictic Gestures in Computer Mediated Communication'. *Intl. Journal of Human-Computer Studies,* 69 (6), 362-374 doi:10.1016/j.ijhcs.2011.01.002

Walther, J. B., & Parks, M. 2002. 'Cues filtered out, cues filtered in: Computer Mediated Communication and Relationships'. Knapp, M. I. & Daly, J. A. *Handbook of interpersonal communication,* 529-563. 3rd edition. Thousand Oaks, CA: Sage Publications.

Wickey, A. & Alem, L. 2007. 'Analysis of Hand Gestures in Remote Collaboration: Some Design Recommendations'. *ozCHI 2007 Proceedings*, 87-93.

Winnicott, D. W. 1971/1990. *Playing and Reality*. New York: Routledge. (First published in 1971).

Winograd, T. & Flores, F. 1986. *Understanding Computers and Cognition: A New Foundation for Design.* Ablex Publishing Corp, (Norwood).

Wright, G. H. von 1998. *In the shadow of Descartes: essays in the philosophy of mind.* Kluwer Academic Publishers.

Wu, Y. C. & Coulson, S. 2007.' How iconic gestures enhance communication: An ERP stud'. *Brain and Language*, 101, 234–245.

Zacks, J M. & Tversky, B. 2001. 'Event structure in perception and conception'. *Psychological Bulletin*, 127(1), 3-21. Jan 2001.

Appendices

Appendix 1 Topic related papers by the author

Lievonen, M. 2009. Attention in Focus: Manipulating Attention in Joint Efforts over Distance. *IIAS-Transactions on Systems Research and Cybernetics*, 9 (1), 11-15

Lievonen, M. 2009. Where is the Focus of Tool Development in Design Education? *Acta Systemica*, 9 (1), 1-5

Lievonen, M. & Rosenberg, D. 2010. An Integrated Approach to Workplace Design: Complexity of Deixis in Video-Mediated Collaboration. (CD-) *Proceedings of Nordic Association of Architectural Research Annual Conference: From Gesamtkunstwerk to Complexity - Architecture in All Scales (NAAR 2010)*. Tampere, Finland, April 22-24, 2010

Lievonen, M., Rosenberg, D., Doerner, R., Kuehn, G. & Walkowski, S. 2009. Augmented Reality as Means for Creating Shared Understanding. Norros, L., Koskinen, H., Salo, L & Savioja, P. eds. *ECCE 2009 - European Conference on Cognitive Ergonomics: Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments*, 161-168. VTT Symposium 258, Helsinki. Available at http://portal.acm.org/citation.cfm?id=1690531

Lievonen, M. 2007. Architecture as Complex Systems: Using Tools & Organising Daily Life. Leong, Y-T. & Lasker, G.E. eds. *Architecture, Engineering and Construction of Built Environments: Proceedings of the Special Focus Symposium on Systems Research in the Architecture, Engineering and Construction (AEC) of Built Environments*, Vol. 2, 15-19. The International Institute for Advanced Studies in Systems Research and Cybernetics

Lievonen, M. 2006. Spaces of Collaboration: Place Taking for Sense Making. *APROS 11: Asia-Pacific Researchers in Organisation Studies 11th International Colloquium*, 268-274. Melbourne: Asia-Pacific Researchers in Organisation Studies, 2006. Available at http://search.informit.com.au/documentSummary;dn=308002422111211;res=IELBUS

Lievonen, M. 2005. Workplace 'Deconstructed'? Leong, Y-T. & Lasker, G.E. eds. *Architecture, Engineering and Construction of Built Environments: Proceedings of the Special Focus Symposium on Systems Research in the Architecture, Engineering and Construction (AEC) of Built Environments. Inter-Symp-2005*, 17th International Conference on Systems Research, Informatics and Cybernetics. August 1-7, 2005, Baden-Baden, Germany, 44-48. The International Institute for Advanced Studies in Systems Research and Cybernetics

Lievonen, M. 2004. Regulating Interaction Space: Layers of Communication and Interaction in Workplace Design. Salavrakos, I. D. ed. *Aspects of Globalisation, Regionalisation and Business*, 179-190. Athens: Athens Institute for Education and Research

Lievonen, M. 2003. From City Walls to Digital Boundaries – Studying Interaction in Hybrid Environment. Rimmer J & Zayas, B. eds. *Human Centred Technology Workshop 2003 Proceedings: Square Pegs in Round Holes? The Relationship between Empirical Research and Theoretical Frameworks*. University of Sussex Cognitive Science Research Papers CSRP 567, 12-14.

Appendix 2 Confirmation of IS-VIT project contributions

Confirmation

to Mirja Lievonen's PhD thesis

This is to confirm that our individual contributions, relevant to her PhD, were the following in the IS-VIT Project:

Guido Kuehn's contribution in IS-VIT Project was to design and implement the 1st test-bed application for the usability tests.

Mirja Lievonen's contribution was

- to design the scenarios for video-mediated pointing
- to design general conditions for the 1st test bed application by Guido Kuehn
- a seamless collaboration with Guido Kuehn (application designer of the 1st test bed) in developing the concept
- the design of the usability test:
  - test setting
  - test tasks
  - method of analysis to validate the concept from communication point of view.

The tests were run jointly.

The sessions of analysing videorecorded data were run jointly, each of us focusing on our individual research questions.

Mirja Lievonen's focused on the adoption of the system, adaptation to the system as a frame of interpretation, and communicative feasibility of the concept (system's capacity to support users in creating shared understanding).

Ingelheim am Rhein / Germany, 16th of March, 2011

Guido Kuehn

Appendix 3 Sample of videoethnography

The interplay of communication modes in co-action

'Closing the session'

Video clip, 31 moves, 43 sec.

People clockwise from the left: A, B, C, D, E, F, G, H (only the feet of person I seen in the picture)

Picture 1 Symptoms of ending: a participant folding his notes (at the right)…

Picture 2 … and stroking his hair = more motor activity (restlessness?) appearing.

Picture 3 Man on the left passing a business card to the lady on his left, the man in the back straightening his posture.

Picture 4 Feet moving to 'ready steady' under the table (man in the corner)

Picture 5 Smiles and laugher – shared politeness / relief?

Picture 6 Perfect timing as if carefully rehearsed.

Appendix 4 Test tasks and instructions

INSTRUCTIONS FOR TEST GROUP (on the 1st test bed application)

The chair role ('David') got instructions relevant to his role, the participants respectively those for the participants only. Instructions were on separate pages. The instruction pages were slightly refined in the course of the tests.

INSTRUCTIONS FOR

EXERCISE 1

For the chair/lecturer ('David'):

You are at the beginning of a videoconferencing session.

- Please open the videoconferencing session by welcoming people and introducing them to each other
- While doing that, make use of AR as you find it best for indicative, clarifying and informative purposes.

Make use of AR as best you can.

For the participants:

You are at the beginning of a videoconferencing session.

- The chair will open the session.
- You are one of the participants in the beginning of a meeting.
- Participate as appropriate to you in such a situation.

After that the exercise is over.

EXERCISE 2

For David:

This is a lecture situation.

- Give a presentation/lecture to the audience. (For instance a 1-2 minutes' talk about the weather in your country).
- Finish it then and ask if the participants have any questions about it.
- Give answers to their questions.

For the participants:

This is a lecture session.

- Participate by making a couple of questions after the presentation you will hear.
- Indicate whether you are happy with the replies given.

After that the exercise is over.

EXERCISE 3 (version a, which was later modified into version b)

For David:

You are a sightseeing tour in London. The remote participants may not be familiar with the area. In the remote location, you can see a map of London. Try to explain for the remote participants something about London.

- Ask them to show the map to the camera and indicate the scope of the City of London in 16th century. (You have just lost the map of yours, but you know about how wide it was.)
- Point then to a few tourist attractions, for instance St Paul and Tower of London.
- Finish your account of the City of London and move to another area, to the City of Westminster.
- After that show on map Bedford House and British Museum, which you can spot one the map behind the participants. Tell something about the surroundings of Bedford House.

For remote participants:

You are discussing a sight-seeing in London, but you are not familiar with the city. The chair of the session tries to give you a brief outline.

- He may need your help in that, so please participate by helping him in his task.

After that the exercise is over.

EXERCISE 4

For David:

You are discussing people's experiences on A30 between Englefield Green and Egham from the point of view of pedestrians and cyclists. The remote partners have the map of the road.

- Find in a group discussion three problem points on A30 from the point of view of pedestrians/cyclists.
- Discuss them as a group, using AR functions available for e.g. indicating, describing and pooling material for decision making.
- After a short discussion, make a group decision which of three areas is the worst of them.

For remote participants:

You are discussing A30 between Englefield Green and Egham. You have the map, but the other group has not got one.

- The task is to define in a group discussion three problem points on A30 (from the point of view of pedestrians/cyclists).
- Discuss them as a group and make use of the map where possible.
- The chair may ask your help during the discussion.
- After a short discussion, decide as a group which of the three areas is the worst of them.

After that the exercise is over.

EXERCISE 5

For David:

- Suggest that you have the next meeting at Kingswood.
- Get the feedback from the remote audience.

- You have left last time you were in the remote location a map there. You can see some items in the remote location. Help them to find the map you left.
- Ask Sam to show the map and help her to find where Royal Holloway and Kingswood are on the map.
- Then show on the map and describe the way from Royal Holloway to Kingswood.

For remote participants:

The place of next meeting is suggested. You don't know where the location is.

- Could Sam tell the chair she does not know the place.
- Help the chair to explain the place by doing what he asks you to do to find out the way to the place.

After that the exercise is over.

2ND ROUND OF TESTS (on the 2nd test bed)

The aims for the second round of tests:

Aims of the tasks 1-5:

1. To see
    a. how two-way pointing works in the introduction, in particular when people at the other end 'intervene' by pointing.
    b. Precision of pointing
    c. How the face detection works (switching the pointer representation from arrow to spotlight).
2. To simulate
    a. the question hour after the lecture;
    b. interactive pointing, when the remote audience wants the lecturer elaborate something which is represented in a document at his/her site.
3. To see
    a. how negotiation works when a tacit document is only at one end;
    b. whether people who are collocated with the document prefer to point at the 'real' document or at its representation on the screen.
    c. There may be also some changes made, so someone has to mark them on the 'real' document.
4. This is a mirror image of exercise 3:
    a. now the tacit document is at the opposite end, and the similar choice of pointing at the 'real' document or its representation on the screen is in question.
5. To see
    a. whether people prefer the tacit document or the representation when working out the way to a destination. There are on both sides the same tacit documents available, but there is only one person who knows the way (next to the chair person).

TASKS FOR TEST GROUP (on the 2nd test bed application)

EXERCISE 1

For the participants:

You are at the beginning of a videoconferencing session.

- The chair (Sam) will open the session.
- You are one of the participants introduced.
- Participate as appropriate to you in such a situation.

For a person (maybe Pat?) who holds the pointer at the remote site (seen from the person who runs the introduction)

- Participate in the introduction by making use of the pointer. For instance: mention that you know two persons from the other end before this meeting (indicate them by pointing), but the third one you may only have seen at a conference last year in London. Ask him/her whether s/he was there.

After that the exercise is over.

EXERCISE 2

For the participants:

This is a lecture (by Alex talking about his/her home country)

- Participate by making a couple of questions after the presentation.
- You see a picture behind the lecturer. Point at something in it and ask the lecturer to tell more about it (whatever you are pointing at).
- Indicate whether you are happy with the replies given.

After that the exercise is over.

EXERCISE 3

For the participants:

You are discussing a floor plan. Chris will introduce the picture and request you to choose one of the rooms to be your office.

- Indicate which of the rooms you prefer.
- Find something you would like to change in the floor plan. Ask for the floor as you do in a meeting (Chris is not in the role of the chair) and point it out and explain what you would like to change.

Once the suggested changes have been made, the exercise is over.

EXERCISE 4

For the participants:

One of the groups has a map of the campus area, whereas the other does not have one. Instead, they may have some pictures behind them.

- Pat asks the group to tell their experiences as pedestrians and cyclists on A30 between Englefield Green and Egham.
- Identify a few problem points along A30, and indicate them on the map.
- After a short discussion, make a group decision which of the problem points you regard as the biggest one.

After that the exercise is over.

EXERCISE 5

For the participants:

This session is about deciding the location of the next meeting. (Eddie is the chair now.)
The place of the next meeting is suggested by Eddie.

- Sam has to tell s/he does not know the place.
- The chair and others will explain the way indicating it on the map.
- After that repeat it by indicating on the map that you remember where to go and where to turn.

After that the exercise is over.

Appendix 5 Questions for feedback discussion

Guiding questions for the discussion after the usability tests
ML / Nov 2007

1 – pointing and giving additional information

What was it like to use the pointer?
What were the problems of using it?
Easy to use? Did it feel natural?
Does pointer support communication?
Does information mode support communication?
(Does snapshot mode support communication?)
(>>> Later: Did it support in control and coordination of the meeting?)

- **remote participants:**

Did it feel surprising?
Did it feel comfortable to be highlighted?
Did it make easier to follow whom the lecturer was addressing?
(>>Later: Did it support in creating the common ground?)

2 - using pointer to indicate the addressee

Was it easy to use? If not, why?
Was it informative? In which way?
Did it make easier to communicate with the addressee?
Did it support turn taking?
(>> Later: Did it support maintaining a common ground? How?)

3 - indicating the scope and the focus of discussion and moving from topic to topic

Is it easy to use?
Did it support indicating the coverage of the topic?
Did it support to indicate the focus of the attention?
Did it support the transitions during the conversation?
Did it support in building the common ground? How?
Did it support coordination and control? How?

4 - indicating points and piling information for decision making

Easy to use?
Did it support in articulating the scope and the focus of the discussion?
Did it help decision making?
Did it support maintaining the common ground?
Did colour support communication?
Did sticker mode support decision making?
Did snapshot support preparing decision making and voting?

5 - using markers on a tacit object

The differences in pointing by finger and pointing by Wii Remote:
Speed? Precision? Informativeness? Social presence? Gestural presence?
Did the pointer support communicating a tacit object in the remote location?
Did information mode support communicating a tacit object in the remote location?
Did markers help maintaining common ground during the conversation?

Appendix 6 Post-test questionnaire

AR-USABILITY TEST
Question form for the participants to be filled before the feedback discussion
GK & ML 19.06.2007

Please tick the following boxes and answer briefly to the questions:

1. Sex: woman ☐ man ☐

2. Age: 20-35 ☐ 35-55 ☐ 55-65 ☐

3. Have you been videoconferencing before? Yes ☐ No ☐

4. If yes, how often? 1-5 times ☐ 5-20 times ☐ >20 times ☐

5. Does AR make any difference in videoconferencing?

None	1
Hardly any difference	2
Some difference	3
Clear difference	4
Great difference	5

6. Arrange the pointer options used in the test in your favourite order, regarding how well they mediated the reference of pointing (1=the best, 2=the second best, 3=the third, 4=the least good for the purpose of pointing)

highlight circle	□
red ring	□
yellow dot	□
arrow	□

7. What is the difference for the better?

1._____
2._____
3._____

8. What is the difference for the worse?

1._____
2._____
3._____

Thank you for your contribution.

Appendix 7 CASE 2 - Methods for specifying user requirements

When a learning interface is in question, the two main perspectives are those of teacher and student. Yet, the role of AV personnel is accentuated when video-mediated learning interface is designed.

In the following, the method is described through teacher perspective. As specifying students' requirements follow the same methods (e.g. observation, questionnaires, focus groups and possible simulations with test users), it has been omitted in the thesis[121].

Teacher as user perspective

We conducted initial discussions with representatives of teaching staff in order to get an overall picture of teaching culture at each university. The discussions followed, in a flexible way, a semi-structured template we prepared in advance for a group and one-to-one conversation.

The aim of the discussions was to identify and classify issues of relevance for the design of a virtual classroom environment that would be satisfying to both staff and students, and therefore candidates for what we should focus on in more detail in Phase 2.

We also wanted to capture particular concerns of teaching staff regarding the use of ICT in teaching, as well as possible differences in familiarity and personal attitudes towards using advanced technology. We summarized the notes of the meetings and requested comments from our informants.

Another method applied in capturing user requirements from the teaching staff point of view was ethnographic observation of teachers 'in action on site'. Again, the aim was to identify topics and issues we should focus on in more detail in Phase 2: we wanted to see in which ways they deliver their teaching in practice, such as

- how they use space and resources in general,
- what kinds of interaction patterns they rely on.

The main aim was to understand why they were doing something in a particular way: what was the motivation / justification of a particular choice. We made notes while observing, and provided that we got their consent for doing so, shot pictures and short videos to capture movements as a flow.

The notes were then organised into a verbal description. It was followed by an initial interpretation and a list of possible design implications based on individual teaching practices. The description with an initial interpretation was then sent to the observed teacher for comments and possible amendments. A few of them provided a significant contribution, helping us to clarify and specify their individual requirements.

In further discussions with them, the description was refined, interpretations amended, and design ideas and implications discussed. The scrutinized version therefore contributed to our findings in several ways.

We also organised sessions with a few teachers, to view their pictures and videos of episodes shot during observation, to discuss them and thereby clarify our understanding of the patterns and strategies employed during the lecture.

[121] Another Swan researcher investigated student experience

The one-to-one session between the researcher and the teacher was the way to increase both sides' understanding of the data. The pictures provide rich material for video-ethnographic analysis of the communicative events.

In addition, they serve as a helpful interface object between the researcher and the teacher to articulate and gain better understanding of individual user requirements. In the next phase, we intend to include them in an active role as an interface object to support inter-professional communication and training.

We had a chance to observe 8 teachers during the Autumn term. The distribution of the observed teachers is uneven across the three institutions, because fewer opportunities were available than expected. We would therefore like to carry out further observations in Phase 2 in order to balance the record.
Interview template on the following page.

Roleplay with AV staff

Video-mediated learning interface poses AV staff in a challenging position: in order to provide adequate technical support, they have to understand what is *relevant* for communication between teacher and student.

Role play (a kind of cognitive walk through) was used as a method to map the technical constraints for design. The senior member of AV staff was positioned behind the lectern, as if a lecturer. That was the way to help him figure out what the teacher actually needs

- to see
- to hear
- to do

in order to communicate the topic to the students (local and remote-site students). (From the student's point of view, the questions are the same from the opposite angle: what s/he needs to see / to hear / to do in order to *understand* the topic explained).

The researcher positioned herself in the auditorium. By providing a number of questions, she guided the AV professional to think about facilitation from teacher's and student's point of view. The challenge for design is to provide, through technical connections, spatial layout and positioning of the equipment, effective learning interface.

Interview Template

SWAN Video conferencing usability / draft refined 29.09.2008 / ML

Interviewing teaching staff at […]

For the interviews at […] , date and time:_____
Interviewer
Names of the interviewees

Interview template

The interview will take an hour (max)
In case you would like to make some notes beforehand, we are going to discuss your teaching according to the following lines.

Our aim is (1) to get an overall picture of how your teaching is delivered in general.

We will view it as

- an 'interface' between instructor and student, seen from your perspective.

That provides some background information for us before your teaching session over a video link. We would like to run another interview after that (in November).

Our aim is (2) then

- to gain understanding of the spatial arrangements of learning sessions over a video link, in order to support the students' learning experience in a virtual classroom
  (>>underlying design principles for a virtual classroom).

What do you teach and how do you do it?

1. Description of your way of teaching

1. What is the topic of the course / **lecture** / lab work / fieldwork?
2. What tools are used (tacit objects, types of artifacts)
3. What interactions are typical for the teaching session? (with items, tools, people)
4. What is the spatial layout of teaching session?
5. What 'choreography' is typical of delivering your teaching / interactions?
6. What is your underlying pedagogical model / thought?

Describe a typical case of your

- **lecturing**
- (lab work
- fieldwork)

Describe the whole range of points1-6.

2. Problems encountered

1. Any typical problems in your teaching regarding
   - settings /
   - methods /
   - patterns of interaction?

3. Design ideas

1. How would you alleviate existing problems?
2. Any suggestions for the design of virtual classroom? (Including the physical settings and technology mediation, i.e. 'hybrid' space).

ASPECTS

A Instructor in focus:

What does the instructor want to demonstrate (intention = the students to focus on?)
What has to be in focus?
In which way – e.g. from which angle, how close, etc.
What kind of spatial arrangement would support such presentation?
Something else relevant:

B Student in focus:

What helps in drawing the student's attention to the topic issue

- visually
- acoustically
- layout-wise?

How long 'passages' or 'chunks' (in time)?
Something else relevant:

C Subject in focus:

What would be the way to present the topic in an informative way?
In an interesting way?

D Context in focus:

What size of group / lecture hall would be optimal / is available?
What shape is the lecture hall / place where teaching is delivered?
What layout there?
What equipment available?
Other – such as lighting, air conditioning, acoustics, etc.
Atmosphere in general – what is important?

E Equipment in focus:
What tools required / used?
What size?
Spatial arrangement of tools
What is important?

F Something else that should be taken into account

Thank you very muchfor your contribution.

Appendix 8 Communicative situation summarized from a spatial design perspective

Table 15 Summary: Communicative situation in spatial design

COMMUNICATIVE SITUATION (in spatial terms)	Segmentation	Structural and functional role	Features	Design questions	Design implications	Design instruction to enable communication / collaboration
Physical space	Physical domain containing set of spatial objects: Non-human Human	Infrastructure as physical grid Medium as carrier of signal (light, sound, smell, temperature, pressure)	Spatial layout of objects as spatial entities >> Individual positions in relation to one another	Presence in the situation = range of attendance (i.e. synchronity as overlap attendance, asynchronity as 0 degree overlap)	Base (grid) for individual perspectives in collaboratio	To provide natural or technology-mediated connection between spatial perspectives.
Perceptual space ('I-here-now')	Sensual domain containing of a set of sensor modalities -Audial -Visual -Haptic -Olfactory -Gustatory	Infrastructure as consensual field -Operational range of a different modalities -Range of involved sensory modalities	Origo of observation = individual perspective opening from a particular spatial position onto the world Range of sensory field: (technically) operational range of particular sensory modalities relative to the capacity of observer's sensory apparatus	Degree of availability of information concerning an object = object contour available re point of view -Which modal channel in question? -Degree of mutuality as overlap of sensory fields?	Mutuality to be created in terms of -operational range of particular modal channel (natural or artificial), relativ to -sending and reception of signals, and -purpose of collaboration	To open modal connection between participating perspectives, through a set of channels that is required /relevant to co-operation.
Social space (or: social-conceptual space)	Cultural domain containing set of common frames of human behaviour In terms of - Built environment (buildings and technology) -Institutional settings -Patterns of social conduct -Language use -Role allocation -Etiquette and politeness	Cultural tradition as shared infrastructure -Operational range of symbolic communication and learning in the shared cultural background (=overlap of backgrounds)	Accumulation of shared knowledge Scope of conventions Patterns of social interaction and communication Range of human knowledge	Degree of shared cultural background -Degree of familiarity with (=knowledge/ ignorance of) mutual cultural traditions	Reciprocity to be created in terms of -social conduct in real life, i.e. providing -control of connection to participants in their -Reciprocate roles (stakeholders)	To adjust the (operational) frame according to the participant perspectives and topic of co-operation.

Cognitive space (or: individual-conceptual space)	Personal world view Mental frame	Operational range of an individual in social situations, -in different roles -contribution to language use -contribution to joint activities	Scope of individual knowledge: worldview as a (trajectory/chain/ stratification/ history) layer of prior encounters and experiences of interaction with the environment	Degree of mutual familiarity, based on prior encounters and communication / relevant information	Control to be provided to regulate the scope of personal privacy in technology mediated settings.	To provide controls to regulate the connection between participating perspectives in terms of mutual / joint privacy, and external involvement.
Orientation	Technical frame of reference, as scope and focus in search of information Practical scope and focus of interest (=motivation)	Cognitive control for -search of information and -orientation in space and time	Attention = observer in relation to the object of observation	Initiation and acceptance of contact in order to activate mutual connection and thereby, co-operative activity	Control to focus on to be provided to enable relevant objects within a topic frame.	To provide controls to manipulate mutual attention to the object of reference.
State of cognition	Situational mental content -Cognitive aspect -emotional aspect -conscious vs. subconscious aspect	Relating new information to prior knowledge structure (as frame of reference) -updating knowledge through sense making -updating values	Identification of/with what is going on, as subjective experience of the situation	Degree of sharedness with prior experiences -Familiarity / unfamiliarity	Interpretation of the situation by a participant, to assess motivation / obligation to co-operate	(Personal participation, through individual motivation (obligation), skills and capacities to co-operative activity).
Verbal communication	Domain of discourse -topic a set of languages for -informing others -being informed by others	System of signs as infrastructure, symbolic tool of -communication -human orientation and adaptation -personal expression	Language use: Expression of views Statements Questions Suggestions Decisions	Role allocation through shared topic and contribution through turn taking		(Personal contribution, in relevant role, through turn taking in co-operative activity).
Deictic gesture	Domain of spatial behaviour and non-verbal communication	Tool for object and attention manipulation and deictic referencing	Informing through nonverbal mode of communication Nonverbal expression of views, attitudes and procedural cues	Tools of mutual attention manipulation	To enable visual availability and common ground plus relevant pointing functionality	(Mutual informing and control of communication through verbal and nonverbal means such as subtle body language).
Gaze direction		Tool for attention manipulation and deictic referencing		Mutual visual availability	To enable mutual visual availability	
Facial expressions		Tool for expressing attitude and emotion		Mutual visual availability	To enable mutual visual availability	