Annotations of Maps in Collaborative Work at a Distance

THÈSE N^{O} 4116 (2008)

PRÉSENTÉE LE 20 JUIN 2008

INFORMATIQUE ET COMMUNICATIONS
RECHERCHE ET APPUI POUR LA FORMATION ET SES TECHNOLOGIES
SECTION D'INFORMATIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mauro CHERUBINI

Master of Arts, St Patrick's College, Dublin City University, Irlande
et de nationalité italienne

acceptée sur proposition du jury:

Prof. P. Fua, président du jury
Prof. P. Dillenbourg , directeur de thèse
Prof. S. Bagnara, rapporteur
Prof. D. Gergle, rapporteur
Dr D. Gillet, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL
2008

Abstract

This thesis inquires how map annotations can be used to sustain remote collaboration. Maps condense the interplay of space and communication, solving linguistic references by linking conversational content to the actual places to which it refers. This is a mechanism people are accustomed to. When we are face-to-face, we can point to things around us. However, at a distance, we need to recreate a context that can help disambiguate what we mean. A map can help recreate this context. However other technological solutions are required to allow deictic gestures over a shared map when collaborators are not co-located. This mechanism is here termed Explicit Referencing.

Several systems that allow sharing maps annotations are reviewed critically. A taxonomy is then proposed to compare their features. Two filed experiments were conducted to investigate the production of collaborative annotations of maps with mobile devices, looking for the reasons why people might want to produce these notes and how they might do so. Both studies led to very disappointing results. The reasons for this failure are attributed to the lack of a critical mass of users (social network), the lack of useful content, and limited social awareness. More importantly, the study identified a compelling effect of the way messages were organized in the tested application, which caused participants to refrain from engaging in content-driven explorations and synchronous discussions.

This last qualitative observation was refined in a controlled experiment where remote participants had to solve a problem collaboratively, using chat tools that differed in the way a user could relate an utterance to a shared map. Results indicated that team performance is improved by the Explicit Referencing mechanisms. However, when this is implemented in a way that is detrimental to the linearity of the conversation, resulting in the visual dispersion or scattering of messages, its use has negative consequences for collaborative work at a distance. Additionally, an analysis of the eye movements of the participants over the map helped to ascertain the interplay of deixis and gaze in collaboration. A primary relation was found between the pair's recurrence of eye movements and their task performance.

Finally, this thesis presents an algorithm that detects misunderstandings in collaborative work at a distance. It analyses the movements of collaborators' eyes over the shared map, their utterances containing references to this workspace, and the availability of 'remote' deictic gestures. The algorithm associates the distance between the gazes of the emitter and gazes of the receiver of a message with the probability that the recipient did not understand the message.

Keywords: Computer-Mediated Communication, Computer-Supported Cooperative Work (CSCW), Deictic Gestures, Eye-Tracking, Human-Computer Interaction (HCI), Location-Based Services, Map Annotations, Remote Deixis, Spatial Cognition.

Riassunto

Questa tesi studia come le annotazioni di carte geografiche possano essere utilizzate per favorire il lavoro collaborativo a distanza. Le mappe condensano l'interazione dello spazio e della comunicazione, risolvendo in tal modo i riferimenti linguistici attraverso la congiunzione di contenuto comunicativo ai siti ai quali questo si riferisce. Questo è un meccanismo divenuto familiare. Quando siamo faccia a faccia, possiamo indicare gli oggetti che ci circondano, ma quando interagiamo a distanza, dobbiamo ricreare un contesto nel quale disambiguare le nostre intenzioni comunicative. Una carta geografica può aiutare a ricreare questo contesto, ma altre soluzioni tecnologiche sono necessarie per consentire a dei collaboratori distanti di potersi scambiare dei gesti dimostrativi su una mappa condivisa. Un tale meccanismo viene qui definito come Riferimento Esplicito.

La tesi analizza diversi sistemi che consentono di condividere le annotazioni di mappe. Uno schema tassonomico è quindi proposto per compararne le loro caratteristiche. Due osservazioni sul campo sono state condotte per investigare la produzione di annotazioni collaborative di carte geografiche attraverso dispositivi mobili. Si sono così cercate le ragioni per le quali possibili utilizzatori possano voler produrre tali note e come. Entrambi gli studi hanno condotto a risultati deludenti. Le ragioni di questo insuccesso sono state attribuite alla mancanza di una massa critica di utilizzatori, alla mancanza di contenuto utile e ad una limitata esposizione sociale che l'applicazione consentiva. In particolare, lo studio ha rivelato un effetto di sopraffazione relativamente alla maniera nella quale i messaggi erano organizzati nell'applicazione testata, che ha impedito ai partecipanti di esplorare i messaggi del sistema in base al loro contenuto e di avere discussioni in tempo reale.

Quest'ultima osservazione qualitativa è stata raffinata in un esperimento di laboratorio dove partecipanti, interagendo da stanze differenti, hanno dovuto risolvere un problema in collaborazione, utilizzando strumenti di discussione testuale che differivano nella possibilità offerta da taluni di poter collegare un messaggio alla mappa condivisa. I risultati hanno indicato che le prestazioni di gruppo sono state migliorate grazie al meccanismo di Riferimento Esplicito. In particolare, quando questo è implementato in una maniera che disturba il filo della conversazione, disperdendo visivamente i messaggi, il suo utilizzo ha conseguenze negative per il lavoro collaborativo a distanza. Inoltre, un'analisi dei movimenti oculari dei partecipanti sulla mappa ha consentito di verificare l'interazione dei dimostrativi utilizzati e lo sguardo nella collaborazione. Una relazione fondamentale è stata individuata tra similarità dei movimenti oculari dei partecipanti e le loro prestazioni.

Infine, questa tesi presenta un algoritmo che rileva possibili fraintendimenti nel lavoro collaborativo a distanza. L'algoritmo analizza i movimenti oculari dei collaboratori sulla mappa condivisa, i messaggi scambiati contenenti riferimenti a questo spazio di lavoro, e la produzione di gesti dimostrativi tramite il Riferimento Esplicito. L'algoritmo associa la distanza tra i movimenti oculari del mittente e del ricevente di un messaggio con la probabilità che il ricevente non lo abbia compreso.

Parole chiave: Comunicazione Mediata dall'ordinatore, Lavoro Cooperativo Assistito dall'ordinatore, Gesti Dimostrativi, Movimenti Oculari, Interazione Uomo-Macchina, Annotazione di carte geografiche.

Published works

At the time of submission, several sections of work from this thesis have previously appeared, or are scheduled to appear, in peer-reviewed publications. Below, the full references for these publications are given.

1. Cherubini, M., and Dillenbourg, P. The effects of explicit referencing in distance problem solving over shared maps. In *GROUP'07: ACM 2007 International Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), Association for Computing Machinery, pp. 331–340.
2. Cherubini, M., Nüssli, M.-A., and Dillenbourg, P. Deixis and coupling of partners' eye movements in collaborative work at distance. In *Extended Abstracts of 2007 International ACM Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), Association for Computing Machinery, pp. 9–10.
3. Cherubini, M., Nüssli, M.-A., and Dillenbourg, P. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the International Symposium on Eye Tracking Research & Applications* (ETRA2008) (Savannah, GA, USA, March 26-28 2008), Association for Computing Machinery, ACM Press, pp. 173–180.
4. Cherubini, M., Nüssli, M.-A., and Dillenbourg, P. This is it!: Indicating and looking in collaborative work at a distance. *Journal of Computer-Mediated Communication*, (2008), *in preparation*.
5. Cherubini, M., van der Pol, J., and Dillenbourg, P. Grounding is not shared understanding: Distinguishing grounding at an utterance and knowledge level. In *CONTEXT'05, the Fifth International and Interdisciplinary Conference on Modeling and Using Context* (Paris, France, July 5-8 2005).
6. Cherubini, M., Hong, F., Dillenbourg, P., and Girardin, F. Ubiquitous collaborative annotations of mobile maps: how and why people might want to share geographical notes. In *Proceedings of the 9th International Workshop on Collaborative Editing Systems* (IWCES'07) (Sanibel Island, Florida, USA, November 4 2007).
7. Cherubini, M., Venolia, G., deLine, R., and Ko, A. J. Let's go to the whiteboard: How and why software developers use drawings. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI2007) (San Jose, CA, USA, April 28, May 3 2007), ACM Press, pp. 557–566.

Acknowledgments

First of all, my gratitude goes to my supervisor, Pierre Dillenbourg. He has taught me what it means to be a researcher. He has also encouraged me to strive for excellence while remaining modest. More importantly, he gave my 'hungry' mind the freedom to explore many research directions. Without his advice, teachings, encouragement, and critique of the work, this thesis would not have been possible. I am sure that my future work will owe much to his theoretical imprinting.

Secondly, I would like to thank wholeheartedly my colleagues at CRAFT that helped me in innumerable ways throughout the course of my PhD. They all have strongly contributed to the conception, discussion and critique of the work here presented: Marc-Antoine Nüssli, Nicolas Nova, Fabien Girardin, Patrick Jermann, Mirweis Sangin, Gaëlle Molinari, Fabrice Hong, Khaled Bachour, Guillaume Zufferey, David Brechet, Frédéric Kaplan, and Florence Colomb.

More specifically, many colleagues contributed to the development of ShoutSpace, one of the CAS tools tested in this thesis: Fabien Giradin, Rachna Agarwal, Patrick Jermann, Shuja Parvez, and Zeno Crivelli. Similarly, the same or other colleagues contributed to the development of STAMPS, the ubiquitous CAS used in the field experiments: Fabien Girardin, Siddarth Jain, Shuja Parvez, Jürgen Scheible, Fabrice Hong, and Christophe Perroud.

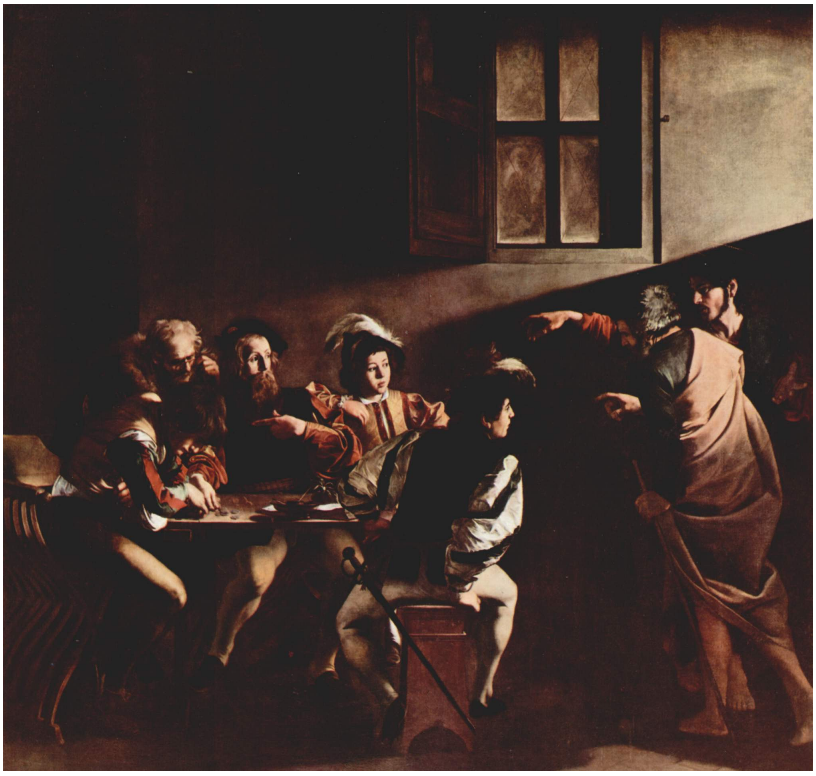
Many thanks also to other friends and colleagues working in other parts of the world who shared ideas, inspired and commented various parts of this work: Jan Chipchase, Elizabeth Churchill, Jean-Baptiste Haué, Cyril Rebetez, Regine Buschauer, Jessica Dehler, Gabriel Fernandez, Carl Gutwin, Pamela J. Ludford, Werner Kuhn, Stefano Mastrogiacomo, Guillaume Raymondon, Francesco Cara, Vincenzo Pallotta, Carlo Ratti, Antonio Scarponi, Lorenzo Viscanti, Andrew J. Ko, David R. Traum, Cristiano Castelfranchi, Pierre Wellner, Jeffrey Huang, Enrico Costanza, and Mark Meagher. Particularly, I am grateful to Jakko van der Pol for co-constructing the linguistic framework used in this thesis; Geoff Underwood, and Daniel C. Richardson for commenting on the eye-tracking analysis presented in this work; Gina Venolia, for inspiring ideas on the beauty of maps; Giles Lane, for sharing many ideas on Location-Based Annotations; and David S. Kirk, for providing me with an example of what a PhD thesis should look like.

I also would like to thank Kamni Gill for proofreading this thesis. Also many thanks to other colleagues at EPFL for being supportive.

I would like to thank all the field trial participants for taking time to use the application, deal with its bugs and sharing their thoughts on location-based annotations (particularly Cyril Rebetez). Also, thanks to all the subjects of the lab experiment for taking time to participate in the study.

Finally, this work was funded by starting founds of CRAFT and by the Kaleidoscope Network of Excellence, European Grant n° 507838 (OFES 03.0558).

A Roberta e Davide,
a voi che essenzialmente siete sostanza
dei giorni e dei sogni miei.
Mauro



The Calling of St Mattew

1599/1600, oil on canvas, cm 322x340 (10′ 7 1/2″ X 11′ 2″), Contarelli Chapel, Church of San Luigi dei Francesi, Rome

Caravaggio represented the event as a nearly silent, dramatic narrative. The sequence of actions before and after this moment can be easily and convincingly re-created. The tax-gatherer Levi (Saint Matthew's name before he became the apostle) was seated at a table with his four assistants, counting the day's proceeds, the group lighted from a source at the upper right of the painting. Christ, His eyes veiled, with His halo the only hint of divinity, enters with Saint Peter. A gesture of His right hand, all the more powerful and compelling because of its languor, summons Levi. Surprised by the intrusion and perhaps dazzled by the sudden light from the just-opened door, Levi draws back and gestures toward himself with his left hand as if to say, "Who, me?", his right hand remaining on the coin he had been counting before Christ's entrance.[1]

[1] From "Caravaggio", by Alfred Moir. See `http://artchive.com/artchive/C/caravaggio/calling_of_st_matthew_text.jpg.html`, last retrieved April 2008.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Space, communication, and collaboration

> *John steps into Mary's office. She is not there. He walks to her desk and picks the APA manual*[1]*, which she uses regularly and that she always keeps in the right-corner of the table. Then he takes a "sticky note", writes his name on it, and fixes it at the place that was previously occupied by the book.*

The spatial environment we inhabit is extremely meaningful to our daily interactions. The anecdote above exemplifies a typical situation in which we take advantage of space to communicate. John did not have to write long explanations, or wait for Mary to return. He did not have to write down the title of the book he was borrowing either, as he was sure that Mary would have remembered which book it was, simply by knowing its position on the desk. This sort of communication is very efficient. It takes advantage of the fact that space offers many affordances to our life. For instance, the place where we live brings to our mind features, goals, and desires, which are completely different from those that we have while at the office. The sticky note used by John is an efficient way of communicating, as it took John little effort to put together its text. It is also a situated fragment of information that the receiver, Mary, finds right where she could have found the object she will be looking for, the missing book. Also, it enables an asynchronous form of communication between John and Mary, as they can converse while not sharing the same space, Mary's office, or the same time (Mary will read the message later).

People naturally take advantage of space to sustain their conversation. Instead of going through complex descriptions, we point to objects as this is an efficient mechanism of resolving the references we use while speaking. When the production and reception of our conversation do not happen at the same time, we record our message in a permanent medium (through a

[1] APA stands for American Psychological Association. The manual is a compendium of formatting and style rules for scientific writing.

written text or audio recording) and we leave clues to our addressee(s) on how to retrieve our words. Signs in a city space communicate directions or positions. They contain only few words or symbols because the place where they are positioned completes their communicative content. Often, this interplay of space and communication happens through maps, as they are easy to reproduce, manipulate, modify and transport to the places where their information might be mostly needed.

While these solutions might lead to satisfactory results in informal communications, collaborative work requires effective methods to coordinate the efforts of the collaborators. Mechanisms which are effective while collaborators are face-to-face might be ineffective, or simply not available, when they are not co-located. For example, it might not be possible for remote pairs to indicate (with a finger) to a certain object while on the phone. In fact, in such situation, the space between the collaborators is not unique and shared anymore. Furthermore, the gestures of one participant might not be visible to the other. Of course, we can think about many possible solutions to overcome this limitation. In this thesis, I will talk about Explicit Referencing (or ER), which I define as the possibility offered by a communication interface to enable its user to enrich a specific message with spatial information. I also term Collaborative Annotation Systems (or CAS), a particular family of remote collaboration and communication tools that implement the ER mechanism through a shared workspace, often a shared map.

The core of this work is therefore to understand the role of CAS tools in mediating communication across distance. **My goal is to understand how Explicit Referencing influences collaborative work at a distance, and to investigate how it impacts a team's cognitive and linguistic processes during collaboration**. Thus, this work approaches the problem from the domain of cognitive psychology and linguistics and yields implications for the design of Computer-Mediated Communication systems and platforms that support cooperative work.

In this thesis, I seek to understand the key design factors of Explicit Referencing and their subsequent impact on remote collaboration. Next, the thesis seeks to translate these results into design implications for CAS. Finally, this work presents a possible application of ER for supporting collaborative work at a distance based on the integration of different communication modalities: the collaborators' deictic gestures, their communication, and the their eye-movements over a shared workspace. To achieve these goals, I designed and conducted three experiments on two custom-made CAS tools that implemented the mechanism of ER in different ways.

This introductory chapter will first present three collaboration scenarios involving the annotations of maps with communication purposes. Then, it will present the research perspective from which the problem will be tackled. Finally, this chapter will conclude with the structure of the thesis and its major contributions.

1.2 Collaborative scenarios

To introduce the theme of this thesis, I will describe three scenarios in which the combination of space, maps, and communication has relevant implications in the support of collaborative work at a distance. Actors and situations described in the following examples are fictional. However, the described practices are not, as they are adapted from existing working situations. The first example describes work practices that are commonly used by teams of archeologists when working on the same site. The second example describes a prototype aviation communication system. The final example concerns the way a control room of a logistic company coordinates the movements of personnel working in the field.

1.2.1 Supporting archaeologists' work on the field

Stefano is part of a team of archeologists working in the excavation of a Byzantine part of Gortina in Crete. The team is composed of 12 members: 6 Italian and 6 Greek researchers. The goal of the project is to find evidence of the inhabitants' practices, rituals and culture. They usually divide the land to be excavated into a grid. Each square is assigned to a researcher who takes care of removing the dust, layer after layer, and cleaning the found artifacts. Sometimes during emergency situations, excavations continues interrupted and different researchers have to alternate on the pits. During each phase of the work, a map of the artifacts found under the soil is hand drawn and annotated with relevant facts by the archaeologist working on that particular sector (see figure 1-1). Relevant aspects of the work are photographed and tagged with a specific code for later retrieval (see figure 1-2). These codes are reported on the maps. All collected materials: maps, notes, photos, audio recordings, references, are cataloged and listed in a wiki archive[2]. These maps are used to share information across team members. They are also used to stimulate discussions on possible interpretations over the collected evidence. Finally, these maps are assembled into a single electronic drawing for documentation and publication purposes.

One night, Stefano dreams of a possible interpretation for a relic that was found the day before by Enrico, another Italian team member: a clay disk with ideograms printed on both sides. He learned of this artifact by re-reading the notes and while talking with his colleagues. The following morning, he prints the updated electronic map of the excavation and he goes on the site, looking for further evidence that could confirm his idea that the disk could describe a religious ritual. First he goes to the place where the disk was found. He looks for the exact point and while looking towards south, he reconstructs in his mind how the city would look in ancient times and whether the sacred mountain was visible from that particular point. The

[2] This example was taken from a real working scenario. The excavation of Gortina was conducted by a team lead by Enrico Zanini of the university of Siena, Italy, during 2006-2007. See `http://www.gortinabizantina.it/`, last retrieved April 2008.

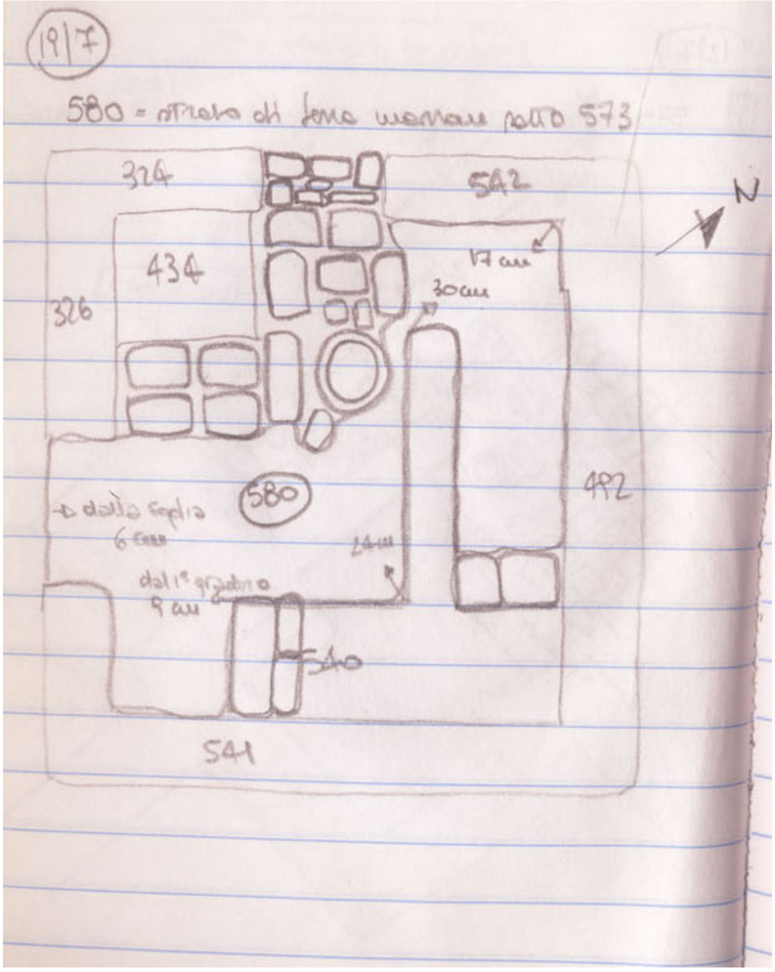


Figure 1-1: Hand-drawn map of one of the excavation pits of Gortina, Crete. The map was subsequently annotated by the archaeologist with relevant facts. Relevant artifacts were photographed and tagged as shown on figure 1-2. This picture was realized by the team of Enrico Zanini and it is reproduced with permission

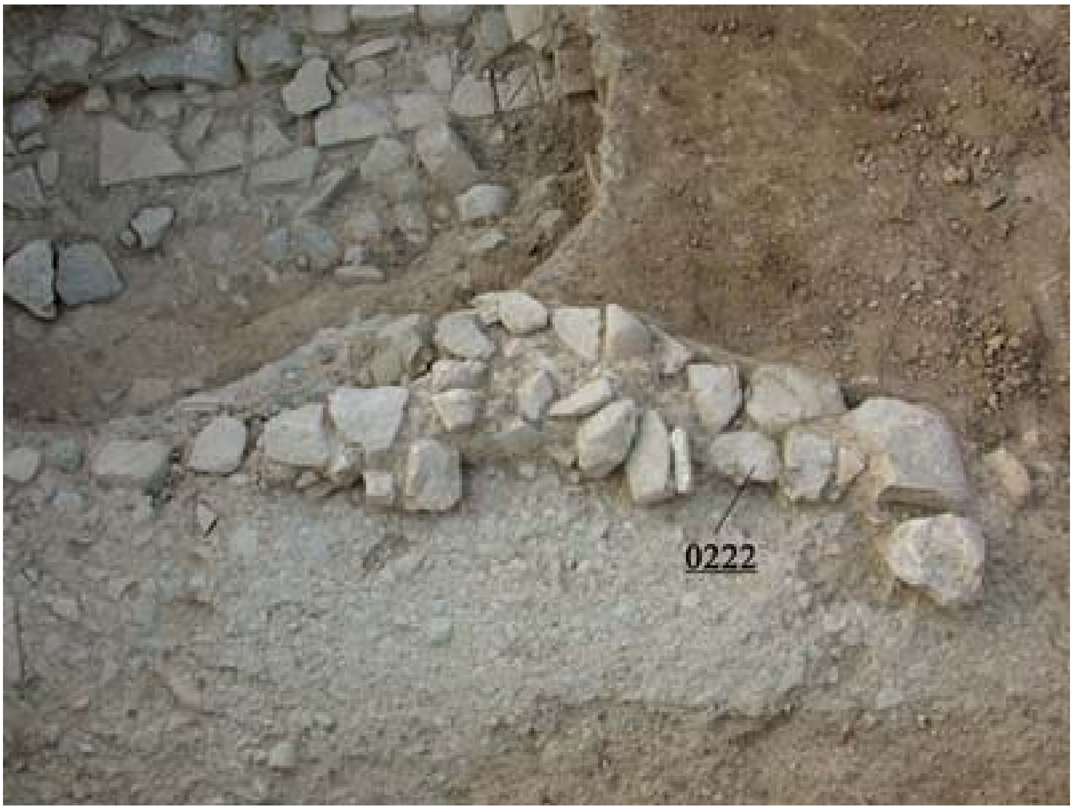


Figure 1-2: Artifact found during the excavation of Gortina. The picture was taken to complement the documentation of the excavation process. The code US 0222 was assigned to the article. This picture was taken by the team of Enrico Zanini and it is reproduced with permission

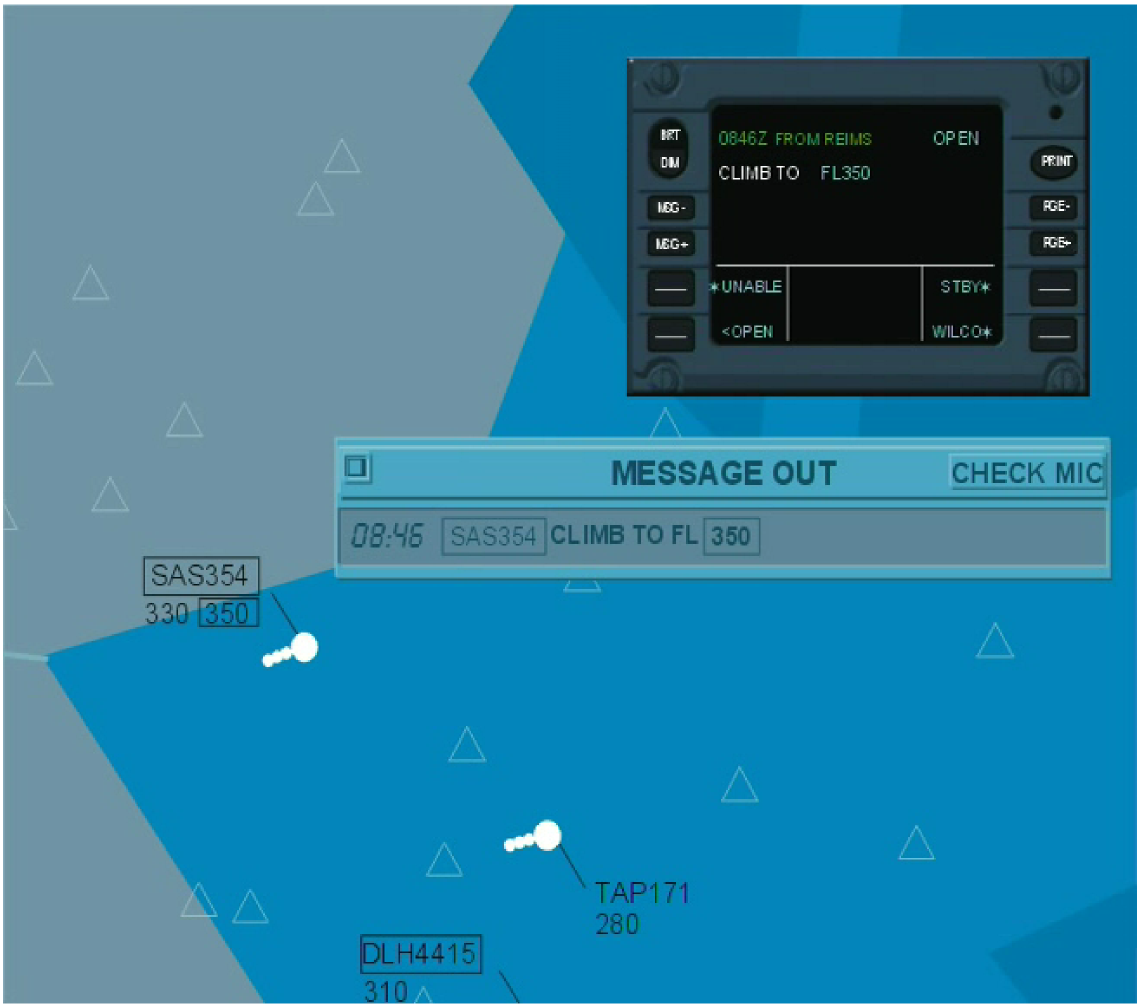
room where the disk was found is at the end of a long, straight corridor, the other end of which is in the direction of the mountain. He scribbles his findings on the map and goes back to the camp. During lunch, Enrico looks at the map that was annotated by Stefano and asks for more information about his idea. After a short discussion, Enrico remembers that the day before, he found a piece of a clay cup with an inscription similar to that of the disk, in a pit situated into another room. This room is at the other end of the corridor Stefano was talking about, thus supporting his thesis.

1.2.2 Communication over a shared map in civil aviation

Ella is a captain working for Scandinavian Airlines[3]. She is flying to Amsterdam on a Airbus 340 which implements the new "Controller Pilot Data Link Communication"[4] (HMI in short) which

[3] This scenario shows an example of textual communication which is situated over a map. However this 'spatialized' communication is mono-directional as only the traffic control sees the messages overlaid over a map.

[4] Link 2000 is a prototype communication systems that is currently in validation phase for the European areospace. The system is supposed to cope with the increasing cluttering of radio frequencies due to the increasing traffic. The system allows the exchange of textual information between the traffic control and the cockpit of incoming aircrafts. See `http://www.eurocontrol.int/link2000/`, last retrieved April 2008.



0846Z FROM REIMS OPEN
CLIMB TO FL350

08:46 SAS354 CLIMB TO FL 350

Figure 1-3: Interface of the HMI system. The traffic controller can click on the radar trace of a plane to activate textual communication with its pilots. The pilot sees the textual message appearing in a dedicated display represented in the top-right corner of the figure (© Eurocontrol 2008)

allows the traffic control to exchange text messages with the airplane. While approaching the airport, her controller sees information of her plane in the radar display. The traffic is congested, so he decides to delay the landing of Ella's plane in favor of a cargo with higher priority. He clicks on the transponder trace of Ella's plane on the radar map and with a combination of clicks, assigns a circular route around the airport to Ella. He sends to her the following message: *"8:46 SAS354 CLIMB TO FL350"* (the first field is the timestamp of the message, while the second is the flight number). Ella sees the incoming message on her radar map and clicking through the interface, she acknowledges the reception of the instructions and the forthcoming execution. Figure 1-3 presents a screen capture of the controller interface.

After 10 minutes, she is still flying on a circular route at flight level 350. She therefore asks the traffic control permission to land. Then, she sends the following message: *"8:56 SAS354 RQS PERMISSION LND"*. The controller sees the incoming message and acknowledges the reception by clicking on Ella's trace on the map and sending the following text: *"8:57 SAS354 PERMISSION TO LND STRP 2 NE"*. Ella acknowledges the message and proceeds to land by aligning the airplane to the assigned second airstrip coming from the North-East direction.

1.2.3 Coordinating the effort of personnel on the field through a control room

Fernando works for a parcel service company called 'Pony Express' in Madrid, Spain. He is riding his bike to a delivery location when he realizes that the address he has is confusing as it is in a neighborhood of closely packed houses without numbers. While in the middle of road works and congested junctions, he sends a short text message with his mobile to the operation room, asking for a correction: *"Cannot find dlvr point. Pls snd exact address. F"*. Miguel, his operator, looks up the address on a map service. He then copies the relevant part of the map that is found by the website, and finally he replies to Fernando's message with a MMS[5] containing a map-tile of the area with a landmark showing the exact delivery point (see an example in figure 1-4). Fernando uses the image to orient himself. First, he takes a reference point, the name of the streed where he stands. Then he reaches the the main junction between "Calle del Tesoro" and "Calle de las Minas", that he can read on the mini-map that he received. From there, he manages to find the address and to deliver the packet.

1.2.4 Commonalities of these different situations

People in these examples share the same need of communicating to others about, or with, resources which are geo-located. In these situations, the map becomes the natural medium through which space is captured, modeled and enriched with communication instances. Instead of going

[5] Multimedia Messaging System. A mobile messaging protocol that allow sending small images, sounds, and text.

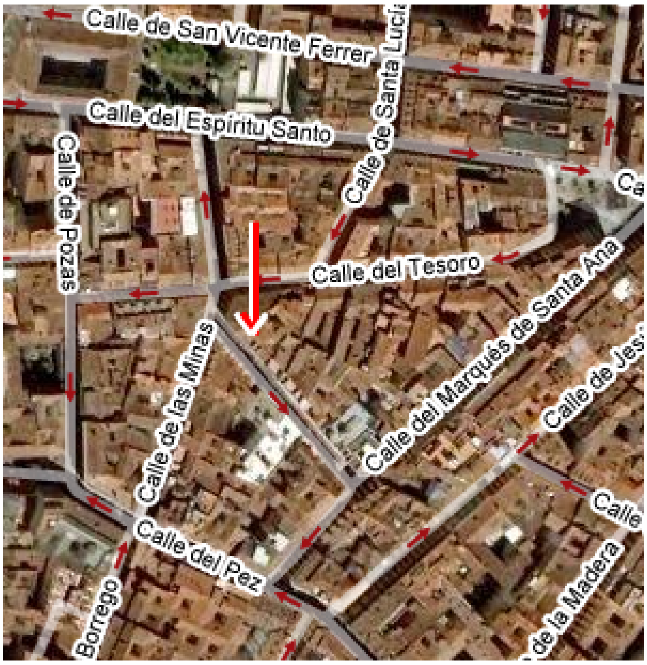


Figure 1-4: Tile of the map that Fernando receives on his mobile. The map was annotated by Miguel with the red arrow to point him on the right delivery point (© GoogleMaps)

through complex textual descriptions of the exact points in space associated with the wanted resources, which might often lead to multiple or erroneous interpretations, people in the examples above choose to share pointers on maps.

In the three examples above, communication bandwidth is reduced to the essential. In the aviation scenario, radio channels are congested by the high number of flights and therefore a text communication channel is used to reduce possible confusion and misunderstandings. In the archeological excavation, bandwidth is reduced, as communication happens asynchronously through the map. Finally, in the delivery logistics example, bandwidth is minimal because the operator cannot point directly on a shared map to the person in the field, so a solution is adopted to overcome this limitation.

1.3 Research focus

To understand why deictic gestures and map annotations are powerful mechanisms for supporting human communication and collaboration, I refer to psycho-linguistic theories which explain how language is situated in space and embodied. Coventry and Garrod (2004a), among others, have explained that the meaning of a sentence cannot be reconstructed entirely from its linguistic nature. Language is polysemous and therefore leads to many possible interpretations. However, once a sentence frames an actual situation, then ambiguities in interpreting the meaning of the sentence tend to dissolve. Additionally, as argued by Lakoff and Johnson (1980), our linguistic knowledge grows out of bodily experiences. Our conceptual system is directly grounded in perception and in the movements of the body. Therefore language understanding is deeply influenced by our own physical nature. These are key arguments that explain why binding communication to space, through maps or gestures, is a natural mechanism that makes communication more efficient and easily interpretable. I will discuss relevant literature in this area in chapter 2.

Understanding how to support collaborative work at a distance is not only related to a deep understanding of which language mechanisms are mostly effective in remote situations, but it is also related to developing an understanding of how different design of communication tools might influence human communication and interaction. Therefore, I refer to cognitive theories which explain how deixis is a fundamental mechanism of collaboration (Daly-Jones et al., 1998). Over the last decades, researchers have attempted in many ways to support remote gestures. Early attempts implemented gestures through the sharing of a video capture of the hands (e.g., J. C. Tang & Minneman, 1991a, Ishii & Kobayashi, 1992). Others have tried to enable gesture at a distance through the use of digital metaphors like sketches or laser pointers (e.g., Bly & Minneman, 1990, Kuzuoka et al., 1994). More specifically, some researchers developed solutions to enable deictic gestures at a distance. An early prototype, called the Telepointer, enabled the transfer of the pointer movements performed on a local machine to the screen of a remote machine, thus enabling a basic pointing mechanism (Greenberg et al., 1996). I will discuss relevant literature in this area in chapter 3.

This research presented positive results of the use of video and sketches to enable remote deixis. However, the majority of these early prototypes used video links between the remote sites. The use of video solutions present many limitations that hampered the ecologies of the remote sites (Luff et al., 2003). I will discuss this issue in details in chapter 3. Furthermore, the adoption of sketch-based solution was supported by the unverified hypothesis that computerized sketches could actually sufficiently communicate hand-gestures at a distance (as discussed by Kirk, 2006). As such, deictic gestures have often been under-considered and few studies have tackled the issue of how to best enable support for deixis in remote collaborative environments. Therefore the general research question that this thesis seeks to answer is: **how can we best design support for deictic gestures for collaborative work at a distance?** The general hypothesis suggested by the literature is that *the availability of an indication mechanism in the communication tool used by remote collaborators could improve performance of collaborative work.* This general question and hypothesis will be refined in chapter 5.

1.4 Thesis overview

The following thesis chapters address the research problem discussed above. This section briefly outlines the content of each of these chapters demonstrating how they evaluate the design of Collaborative Annotation Systems, explore the role of deictic gestures in remote communications and how they progressively build an argument for the design of communication tools offering support for remote deictic gestures.

Chapter 2 — Space, communication and interaction — defines a linguistic framework for understanding the integration of language and space. First, the chapter describes the use of spatial prepositions in language and how their interpretation requires a combination of linguistic and non-linguistic elements provided by the context in which the conversation takes place. The chapter concludes by introducing the linguistic framework that will be used to interpret the experimental evidence provided by this thesis.

Chapter 3 — Deixis in dual spaces — reviews previous research in the area of remote support for deictic gestures. First, the chapter introduces a framework for evaluating Computer Supported Cooperative Work applications, which describes deixis as a basic pragmatic need of communication. The chapter describes the state-of-the-art in remote gesture tools and discusses the evaluatory studies that have been performed with them. The chapter reveals that these studies showed that remote gesture representation and a shared visual workspace are important but also highlight that attempts to support these mechanisms can lead to a fracturing of the interaction between collaborators. Finally, the chapter highlights the interconnections of deictic gestures and gaze awareness.

Chapter 4 — Location-Based Annotations — presents a critical review of CAS. The chapter defines a topology of CAS based on eight different dimensions. Particularly, the chapter profiles three factors that can greatly influence the user experience of CAS tools: the degree of immersion offered, whether they are designed for mobile or fixed use, their organization criteria, whether the messages are ordered by time of by content, and the time-span of the interaction for which they are designed, namely synchronous or asynchronous use. The chapter concludes by presenting some examples of application of Location-Based Annotations.

Chapter 5 — Research methodology — forms hypotheses on the basis of evidence from the literature review and the classification of CAS tools. The chapter explains the rationale for conducting two different kinds of experiments with different annotation systems that implement in different ways the three factors highlighted in chapter 4. The chapter explains the different experimental methodologies that will be used in these two contexts, the tools that I constructed for the experiments, and the statistical methods that I employ to analyze the results of the controlled experiment.

Chapter 6 — Qualitative Observations — presents initial evidence of the impact of different designs of an ubiquitous CAS to the forms of communication that can be spontaneously adopted by its users. The chapter provides answers to the first four research questions, reporting the results of the exploratory studies of field use of STAMPS, one of the applications that I developed for this thesis.

Chapter 7 — The effects of Explicit Referencing in distance problem solving — presents a controlled experiment which demonstrate how Explicit Referencing can improve aspects of performance in remote collaboration tasks when compared to standard communication tools. This experiment examines basic performance metrics, including the time required for the symmetrical positioning of placeholders on a shared map and the score and number of solutions explored in a collaborative task that I designed. The experiment also examines the linguistic process employed by pairs for coordinating their effort in different experimental conditions.

Chapter 8 — Indicating and looking in collaborative work at distance — extends the results presented in chapter 7, by reporting a thorough analysis of the eye-movements of the participant over a shared workspace during the task. This study was conducted with the hypothesis that tools implementing Explicit Referencing would affect the way people look at the shared workspace, and their subsequent performance. The presented cross-recurrence analysis was adapted from a previous study of eye-movements of participant in a listening-comprehension task. The chapter therefore highlights the interconnection of gaze and deixis in supporting collaborative work and favours of the combination of gaze awareness in future implementations of ER. Finally, the chapter reports a qualitative analysis of the eye movements of the participants during the task, which shows that both the emitter and the recipient of a message tend to look at the points of the shared workspace mentioned in the emitted messages.

Chapter 9 — A computational model to detect misunderstandings — extends the hypothesis that was raised in the qualitative analysis of the eye-movements of chapter 8. The chapter presents an algorithm to automatically detect episodes of misunderstanding occurring while participant solve the task tested in the previous chapters. The chapter demonstrates how the integration of multimodality of communication, like eye-movements with deictic gestures and a simple language model, might yield interesting results for supporting remote collaboration.

Chapter 10 — Conclusions — concludes the thesis by summarizing and evaluating the contributions that the thesis make for the design of existing Collaborative Annotation Systems. In synthesis, the chapter highlights the importance of a linear message history as well as that of Explicit Referencing in support of remote collaboration, and also the interrelation of the mechanism of remote deictic gestures with gaze. These findings are considered in the light of the limitations of the conducted studies. Finally, the chapter discusses the implications of the findings presented in this thesis for the design, deployment and development of these technologies, articulating a program of future work to address the raised issues.

1.5 Thesis contributions

Having articulated the structure of the rest of the thesis and how the thesis will address the research area, it is pertinent to conclude this introductory chapter by detailing the overall contributions this thesis makes. The main contribution of this thesis is a through understanding of human factors as they relate to the design and use of Collaborative Annotation Systems. Specific contributions include:

- A through discussion of the requirements of studying Collaborative Annotation Systems and their applications;
- A taxonomy of CAS tools and their communicative uses;
- A set of guidelines for deploying ubiquitous CAS tools, including factors related to the task in which these are used and the participants using them;
- A set of guidelines for designing CAS tools, focusing on the identification of the core criteria for supporting collaborative work;
- An experimental comparison of different communication tools, implementing in different ways, the mechanism of Explicit Referencing and their impact on the occurring collaborative speech patterns;
- An experimental evaluation of the interconnection of deictic gestures and gaze, yielding important implications for the design of CAS systems;
- A prototype algorithm demonstrating the possible detection of episodes of conflict during collaboration and based on the integration of a linguistic model and the analysis of eye-movements.

1.6 Conventions used in the text

Throughout the text, I will often refer to hypothetical situations that can help me illustrate the arguments of this thesis. In these examples, I will often refer to a prototypical user using the personal pronoun 'she'. Whenever the scenario will require two actors, I will use the personal pronoun 'she' for the first actor and the male pronoun 'he' for the second actor, according to the order of appearance in the text of the example.

Additionally in the text, I will report utterances extracted from experiments that can provide evidence for the reported analyses. In these situations, I will type the utterance within double quotes (“ ”) and in *italics*, so that it will be easier to distinguish them from other parts of the text, and from other examples of imaginary situations. Sometimes, single quotes (‘ ’) will be used to indicate metaphors or improper terms.

Italics in the text will be also used for indicating definitions and technical terms that are introduced progressively in the text of the thesis. These terms will be also tracked across the difference chapters in the index of the thesis reported at page **??**.

boldface of text will be used throughout the thesis to indicate specific keywords and to facilitate the recognition of a list of elements described in the same paragraph.

Finally, Sans Serif will be used for the names of research applications and commercial applications that I will analyze or use as examples throughout the text.

Chapter 2

Space, Communication and Interaction

This chapter presents literature developed in the psycho-linguistics domain relevant to this thesis. It introduces the role of space in language understanding and then in language production and use. Next the chapter summarizes relevant work explaining how space influences human interaction. Finally, the chapter introduces the research framework that I will use throughout the rest of the work (see section 2.4).

2.1 Space and language understanding

Finding objects in the world is one of the most basic survival skills required by any living organism. Similarly, describing where objects are, and finding objects based on simple locative descriptions can be regarded as a basic skill for any competent speaker of a language. Understanding how spatial language works poses the challenge of understanding how spatial language is organized within a language (how space is structured by language), but also understanding the connections of language and perceptual representations (how language is structured by space).

In this section, I will first look at the main linguistic approaches to spatial language (section 2.1.1). Then, I will examine the perceptual primitives associated with spatial language and the mapping between these perceptual representations (section 2.1.2). Finally, I will review a number of computational approaches that aimed at mapping spatial language and perceptual representations (section 2.1.3).

2.1.1 Spatial language in context

Many everyday situations require the understanding of spatial descriptions. We use them all the time to locate objects, but more generally, we use them to reason about the world. Expressions such as *below* or *above* can help identify objects in space, but also to reason about floors of a building depicted on a sketch or the relative positions of countries on a map. Although locative expressions are part of standard communication, understanding how these are used and understood by people is an extremely difficult problem. The words that express the location of objects are called spatial prepositions (in *italics* in the examples reported in this chapter).

Spatial prepositions are interesting elements of a language because these are the hardest to learn when a person acquires a foreign language. This is because *languages differ in the way in which they map linguistic terms onto spatial relations* (Coventry & Garrod, 2004b, p. 4). Despite this cross-linguistic variability, each language contains only a few spatial prepositions. There are only between 80 and 100 prepositions in the English language, these are displayed in table 2.1. Landau and Jackendoff (1993) have argued that natural languages only encode a limited number of spatial relations between objects and these have to cover a whole range of possibilities. Language is *polysemous*, meaning that any word may have a range of distinct but related interpretations.

Spatial prepositions are commonly divided into *locative* or *relational* prepositions, describing the location of an object in relation to another (e.g., The apple is *in* the bowl), and *directional* prepositions, describing a change of position (e.g., Paul went *to* the restrooms a few minutes ago). Locative/relational are often further divided into *topological* terms, prepositions usually referring to static relations between objects (including prepositions such as *in*, *on*, and *near*), and *projective* terms, providing information about the direction in which one object is located with reference to another object (including terms such as *in front of*, *to the left of*, and *above*). Projective terms depend upon a particular frame of reference for their interpretation. Levelt (1996), reviewed three perspectives language users can take in mapping spatial relations onto linguistic expressions, the *deictic*, *intrinsic* and *absolute* systems (see figure 2-1):

(1) *Intrinsic*, or object-centered, frames of reference use a coordinate system to specify the position of the located object which is generated with respect to the salient features of the reference object. Positioning expressed in this way is not transitive (e.g., in the case of figure 2-1, if the ape is to the right of the bear and the bear is to the right of the cow, this does not imply that the ape is to the right of the cow).

(2) *Relative*, or viewer-centered/deictic, frames of reference presuppose an egocentric viewpoint distinct from the objects being described. In figure 2-1 this position is taken by the shape of the man in the center.

(3) *Absolute*, or environment-centered, frames of reference are defined with respect to salient

Table 2.1: The prepositions in English (adapted from Landau & Jackendoff, 1993, p. 224). AE = occurs only in American English, SE = occurs only in Scottish English

about	above	across	after
against	along	amid(st)	around
at	atop	behind	below
beneath	beside	between	betwixt
beyond	by	down	from
in	inside	into	Near
nearby	off	on	onto
opposite	out	outside	outwith (SE)
over	past	through	throughout
to	toward	under	underneath
up	upon	via	with
within	without		
Compound prepositions			
far from	in back of (AE)	in between	in front of
in line with	on top of	to the left of	to the right of
to the sight of			
Intransitive prepositions			
afterwards	apart	away	back
downstairs	downward	east	forward
here	inward	left	N-ward (e.g. homeward)
north	onward	outward	right
sideways	south	there	together
upstairs	upward	west	
Nonspatial prepositions			
ago	as	because of	before
despite	during	for	like
of	since	until	

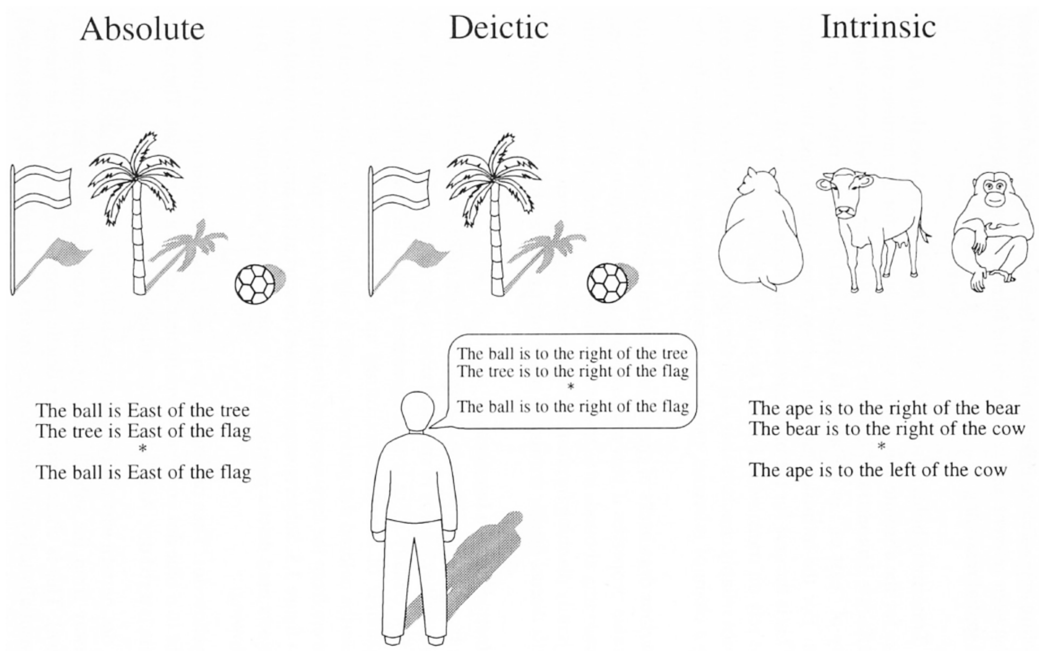

Absolute
Deictic
Intrinsic
The ball is East of the tree
The tree is East of the flag
*
The ball is East of the flag
The ball is to the right of the tree
The tree is to the right of the flag
*
The ball is to the right of the flag
The ape is to the right of the bear
The bear is to the right of the cow
*
The ape is to the left of the cow

Figure 2-1: Frames of reference as noted by Levelt (1996). Transitivity holds for the absolute and deictic system but not for the intrinsic system

features of the environment such as cardinal directions or directions provided by gravity, which are arbitrary.

Simple prepositions like *above* or *below* can also be grouped into a set of expressions, like *on the right of*, that convey spatial relationships (Talmy, 1983). Cognitive linguistics has been regarding spatial concepts as the primary structuring tool for other conceptualised domains, like spatial metaphors (Lakoff & Johnson, 1980): spatial terms used in expression of time (e.g., I'll see you *in* ten minutes), or expression of emotion (e.g., I was feeling *down* yesterday).

Lexical semantic approaches have tried to capture the meaning of lexical items only in terms of other lexical items (see, for example, Lund, Burgess, & Audet, 1996; Landauer & Dumais, 1997). However, Glenberg and Robertson (2000) noted how these approaches, that define words in terms of other words, do not deal with the issue of how meaning maps onto the world. Harnad (1990) also agreed that the meaning of a word can never be figured out without grounding the symbol in something else. In this regard, Coventry and Garrod (2004a) offered an account that descriptions of the spatial world not only refer to the positions of objects in space, but also reflect the knowledge of objects in the world acquired through interacting with that world. Building on earlier work, Coventry and Garrod laid out a *functional geometric framework* for spatial language comprehension and production that incorporated both geometric constraints and extra-geometric constraints derived by humans through their perception and action in the actual space. Therefore, it appears necessary to take into account perception to understand how meaning is inferred from language, as discussed in the next section.

2.1.2 Grounding language in perception

Lakoff (1987) was one of the first cognitive linguists to stress that embodiment is a key element missing from lexical semantic theories, which explain concepts and categories only in terms of fixed boundaries. According to Lakoff, our cognitive structures grow out of bodily experiences. The core of our conceptual systems is directly grounded in perception, movements of the body, and social experiences. Therefore, Lakoff claims that the concepts I possess and the language I produce and understand are fundamentally influenced by our own physical nature. Lakoff and colleagues approached spatial language through the definition of image schemata that were driven by spatial senses (Brugman & Lakoff, 1988). Image schemas are dynamic embodied patterns. They are multi-modal patterns of experience, not simply visual. For example in figure 2-2, one can consider how the dynamic nature of the containment schema is reflected in the various spatial senses of the English word *out*. Out may be used in cases where a clearly defined trajector (TR) leaves a spatially bounded landmark (LM), as in: *Paul went out of the room*. In a prototypical case the landmark is a container. However, out may also be used to indicate those cases where the trajector is a mass that spreads out, effectively expanding the area of the containing landmark: *He poured out the nails*. Finally, out is also often used to describe motion along a linear path where the containing landmark is implied and not defined at all: *He started out for Washington*. Coventry and Garrod (Coventry & Garrod, 2004a) critiqued Lakoff and Brugman's approach to classify geometric relationships between objects in the world which they mapped onto individual image schemata. According to Coventry and Garrod there is not such one-to-one mapping as words can have an infinite number of senses (as noted by Johnson-Laird, 1987).

Understanding spatial language also means understanding the purpose that location serves for the users of that language. Humans acquire this common knowledge through their continuous interaction with the world and learn how to use it in every situation. The knowledge of how a particular object function and moves has an essential survival value. Gibson (1979) described action and movement as basic features of his "ecological" approach to vision.

Contrary to the view that spatial prepositions rely on only schematised properties, the functional geometric framework, proposed by Coventry and Garrod (2004c), argues that objects are what fundamentally influences how one talks about where they are located. According to this model, meaning of situation-specific spatial prepositions is assigned by two types of component: the geometry of the scene and the extra-geometric factors, like how the observed objects can move, their consistency, their weight, and so forth. Objects are associated with their function, which may promote the application of different routines. For example, if we are describing a scene of a person eating spaghetti in a restaurant, depending on how we label the container of

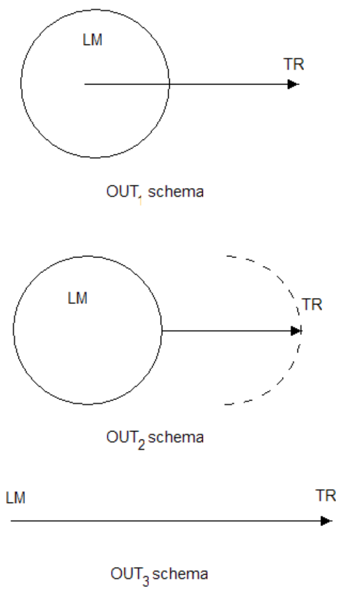


Figure 2-2: Containment Image Schema as applied to the English word *out* (TR, defined trajector; LM bounded landmark), from Johnson, 1987

the spaghetti, different spatial prepositions becomes appropriate. If the reference object is labeled a plate, *on* is judged suitable. Else, if it is labeled a bowl, *in* becomes more appropriate.

2.1.3 Modelling spatial language

During the last few years, researchers have tried to construct models (and computational models) that are able to capture the definition of spatial prepositions as humans employ in natural language. A detailed review of recent trends in this area is given by Mark et al. (1999). In this section, I limit myself to listing a few relevant models that inspired the application of thesis work, as detailed in chapter 9.

Cohn et al. (1997) developed a qualitative geometry of space called the *region connection calculus* (or RCC), which treats "regions of space" as fundamental. Let me consider containment as a prototypical spatial relation for the rest of this discussion. RCC defines spatial relations such as enclosure in terms of just two primitives: *connection* and *convexity*. Connection is a broadly defined relation that covers everything from simple contact to overlap between regions and their identity. Convexity, on the other hand, relates to the presence of an object in a region of interior spaces, defined in relation to what Cohn calls the *convex hull*[1] of the region. According to this model, there are a number of ways in which one object can be represented *in* another object(s). This is reflected in different degrees of enclosure that makes the RCC model suitable to capture

[1] In mathematics, the convex hull or convex envelope for a set of points X in a real vector space V is the minimal convex set containing X.

A	A	A	G	A	A	A
A	A	A	G	A	A	A
A	A	A	G	A	A	A
B	B	B	■	B	B	B
B	B	B	B	B	B	B
B	B	B	B	B	B	B
B	B	B	B	B	B	B

Figure 2-3: Example of a spatial template for *above*, from Coventry & Garrod, 2004c. G = good region, B = bad region, A = acceptable region

a number of different situations where one can say object A is *in* object B. However, the authors did not provide evidence that regions and complex hulls constitute basic perceptual categories. Additionally, even this geometry does not capture the full range of use of spatial prepositions.

An early attempt to develop a model incorporating perception of spatial relations as the enclosure discussed above was that of Ullman (1996), who argued that perceptual processing requires *visual routines*. According to his model, determining whether a object is contained in a complex visual scene might be assisted by a routine like that of starting from a given point and then progressively colouring in the region around the point until a boundary contour is encountered. Ullman argued that the representations delivered by the visual system do not contain any notion of *inside/outside*. Instead, these are elaborated through visual routines. Visual routines serve functional perception, and they are subject to attention control. Therefore, the model does not offer a deterministic output for a given set of preconditions.

In the same year, Logan and Sadler (1996) claimed that spatial templates underlie the comprehension of spatial relations and spatial prepositions. A template is a representation that is centered on a reference object and aligned with the reference frame imposed on it. An example of such as spatial template is represented in figure 2-3. By superimposing a template on a certain scene, it is possible to express an acceptability judgement for a spatial relation (corresponding to that template) applied to all the objects in the scene. According to Logan and Sadler, *spatial indexing* is necessary to establish the correspondence between the spatial template (the symbol) and the perceptual scene (the percept). Once this indexing has been established, it is then possible to judge spatial expressions as appropriate.

More recently, Regier and Carlson (2001) developed the *attention vector sum* model (or AVS). This model takes into account the role of attention in determining a spatial relation and has much the same character as one of Ullman's visual routines. In the model, an attentional beam is focused on a landmark . In particular, the beam is focused on that point of the landmark top that is vertically aligned with the trajector or closest to being so aligned (part a of figure 2-4). Parts of the landmarks near the center of this beam are strongly attended, whereas more distant parts of the landmark receive less attention (part c of figure 2-4). This yields a distribution of attention across the landmark object. The attentional beam radiates out to illuminate different parts of the landmark at different strengths, depending on the distance from the focus. All these vectors are combined to determine a resulting attentional vector (part d of figure 2-4). The authors empirically tested prediction accuracy of the different models by presenting the model with experimental stimuli, recording the model's output, and determining through the regression how well the model output predicted the empirically obtained acceptability rating. The study confirmed the predictions of the AVS model: first, spatial terms ratings are influenced by the proximal and center-of-mass orientations. Second, ratings are sensitive to the *grazing line* (the horizontal line 'scraping' the very top of the landmark). Third, ratings are affected by distance. The model provides a preliminary grounding of linguistic spatial categories in non-linguistic perception: linguistic spatial categories can be explained in terms of underlying structures that are not linguistic in character.

One final work of note in this area is the simulation method developed by Frank and colleagues (2001). They devised a method for simulating human behavior in space using multi-agent systems: mutiple agents acted in an environment that represented the simulated world. They each had a certain base knowledge, processes, and perceptions about the world. This information, which was not necessarily correct, was used to make decisions, to act and to communicate. Other agents could see their actions or 'hear' their communication and use this information together with their perception of the world to make decisions about what actions to take (see figure 2-5). Using this system, they showed how a *map-maker* and a *map-user* agent could make mistakes in the perception and form erroneous beliefs about the environment. Simulated agents observed the environment and formed a set of beliefs about it, which could be incomplete, imprecise or even wrong. The agents produced resulting map-artifacts, which represented their knowledge. The authors then defined *homomorphism*, as the agreement between the agent's beliefs and the actual environment, where objects and operations are set into correspondence. The goal of their work was to have a way of formally checking that the descriptions of formal spatial ontologies were complete, namely that all parts which are used to define a concept were, in turn, defined somewhere else in terms of a very simple set of primitives. This automatic control gave additional confidence that a computational model was logically consistent and that this model correctly reflected the intended behavior.

2.1.4 Summary: space and language understanding

- *Language schematizes space well.* Language is very good at encoding spatial features whereas

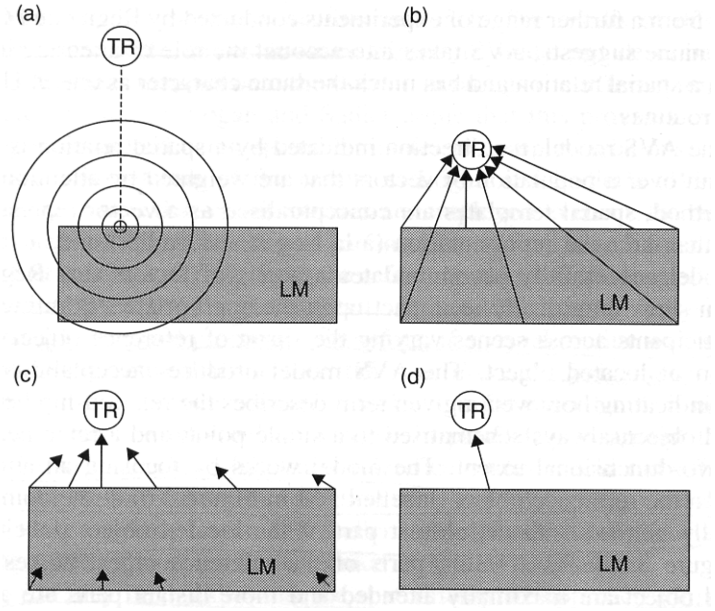


Figure 2-4: The attention vector sum model (TR, defined trajectory; LM bounded landmark), adapted from Regier & Carlson, 2001

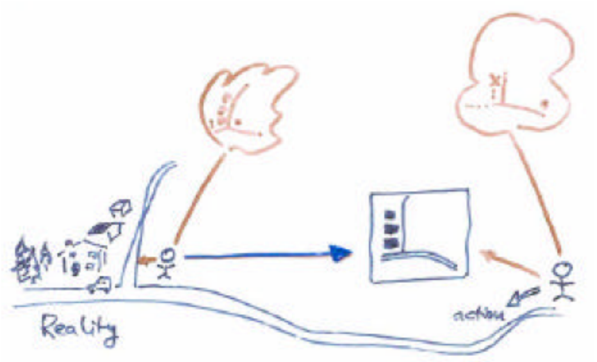


Figure 2-5: An agent producing a map and another agent using a map for navigation, from Frank et al., 2001

is not as efficient at encoding face features. An example of this great ability to schematize space can be given by the English term *across*, where ideally the 'thing' doing the crossing is smaller than the 'thing' being crossed, and it is crossing in a straight path perpendicular to the length of the 'thing' being crossed. Thus schematization entails information reduction, encoding certain features of the scene while ignoring others.

- *Perspectives.* Considering other points of view is essential for a range of cognitive functions and social interactions, from recognizing an object from a novel point of view to navigating an environment in order to understand someone else's position. There are three bases for spatial reference: the viewer, other objects and external sources. These three bases seem to correspond to deictic, intrinsic and extrinsic uses of language. An interesting point is that deictic uses cannot be accounted for by the language alone. They require additional knowledge of the interactional situation in which they are produced. Depending on the complexity of the task, the speaker can decide to take his own perspective, the perspective of the addressee or a neutral perspective, and use a landmark or referent object, on the extrinsic system as a basis for the spatial reference.

- *Embodiment.* Lexical semantics could not explain the entirety of meanings that spatial language can encode. Words can have infinite meanings and the way people manage to communicate successfully is because the symbolism of words is grounded through perception. The concepts I possess and the language I produce and understand are fundamentally influenced by our own physical nature.

- *Functional geometric framework.* Understanding spatial language means also understanding the purpose that location serves for the users of that language. Modelling should therefore take into account geometric constraints as well as extra-geometric constraints. These extra constraints are given by the shared knowledge that people have of physical objects and their interaction with the world.

2.2 Space and language in use

As I have highlighted in the previous section, spatial environments have an objective reality, and language has become rich and flexible in spatial expressions for better encoding its features. While the previous section discussed how spatial language expressions are understood and the way meaning is encoded, here I focus on how people actually use language to describe space. While the studies reported above explained these relations from a theoretical perspective, not always supported by empirical evidence, studies in this section have been developed in an experimental framework and aimed at providing evidence of spatial cognition from user studies. Finally, this thesis focuses on a particular application of spatial language, namely the use of maps and map annotations to sustain human communication and therefore collaboration. Several studies focusing on these artifacts are discussed in the last part of this section.

2.2.1 Spatial perspectives in descriptions

The process of description can be summarized with two steps: an organization and a description process. Environments are organized hierarchically in memory. The features that are larger or whose functions are more significant have priority on the others (McNamara, 1986). Tversky (1977) showed how these consists of landmarks and the approximate spatial relations among them, plus non-spatial extra information. Taylor and Tversky have shown that mental representations are perspective-free. Subjects who learned descriptions of an environment written in a certain perspective, were able to answer inference questions from the perspective they had not read as quickly and accurately as inference questions from the perspective they had read (Taylor & Tversky, 1992a, 1992b, 1996).

Once the information has been organized then it is possible to describe it. Levelt (1996) distinguished two aspects of generating spatial descriptions: *macroplanning* and *microplanning*. In macroplanning we elaborate our communicative intention, selecting information whose expression can be effective in revealing our intentions to a partner in speech. We decide on what to say, linearizing what goes first and next. Levelt argued that ordering follows two principles: the *principle of connectivity* and that of *natural order*. According to the principle of connectivity, each utterance should connect directly with the previous and subsequent utterance. The principle of natural order states that organization depends on the content of the messages (e.g., time to events, or source to goal). Determining perspective is part of microplanning. In microplanning, or "thinking for speaking", we translate the information to be expressed in some kind of "prepositional" format, creating a semantic representation, or message, that can be formulated. Applied to spatial discourse, I can say that macroplanning involves selecting *referents* (the TR in figures 2-4, 2-2), *relata* (the LM in figures 2-4, 2-2), and their *spatial relations* for expression (usually the spatial preposition). Instead, microplanning consists of applying a chosen perspective system that will map spatial directions/relations onto lexical concepts.

The world is multidimensional but speech is linear. It makes sense to choose an order to describe the world linearly. Environments are typically described from *gaze*, *route* or *survey* perspectives, as described by Taylor and Tversky (1996). A gaze perspective describes objects relatives to each other, from an outside viewpoint. A route perspective takes a view from within, and describes landmarks with respect to the changing position of "you", a traveler in the environment, in terms of left, right, front, and back. Finally, a survey perspective takes a view from above and describes landmarks with respect to each other in terms of north, south, east, and west. In *gaze* *tours* noun phrases are usually headed by objects and verbs express states. In *walking tours*, noun phrases are headed by the addressee and the verbs express actions (e.g., if I am giving directions: "you have to go straight ...").

The choice of a perspective and a particular strategy to encode the spatial situation will depend closely on the number of mental transformations required to produce or to understand an utterance (Tversky, 1996). It stands to reason that speakers would avoid cognitively difficult descriptions.

The selection of a certain perspective may depend on how an environment has been experienced by the person giving the description but also on objective features of this environment. For instance, Tversky and colleagues explain how when the number of single paths of the scene is equivalent to that of landmarks, this can encourage a route rather than a survey perspective (*ibidem*, p. 389). The way a describer switches perspective is due to the natural ways s/he captures and experiences the spatial world. Descriptions and representations are dependent on individual preferences and cognitive styles. It is therefore of interest for this discussion to look at how spatial representations and language are combined both at mental level and in actual artifacts, like maps, to support human activities.

2.2.2 Visuospatial Reasoning: Maps, cognitive maps and maps annotations

Reasoning is going beyond the information given (Bruner, 1973). This does not mean necessarily adding new elements, but transforming the given information through deductive reasoning, inferences and judgments.

Visuospatial representations capture visuospatial properties of the world preserving, in all or in part, the spatial structural relations of that information (Johnson-Laird, 1983). The visual features include static properties of objects, such as shape, texture, and color, and their relation to a reference system expressing therefore their comparative distance and directions. These features also include dynamic properties such as direction, path, and style of movement. Visual representations contrast with linguistic representations. Their similarities and differences provide interesting insights (Talmy, 1983).

Shepard and colleagues (e.g., Shepard & Podgorny, 1978; Shepard & Metzler, 1971) conducted several investigations on visuospatial reasoning from a bottom-up perspective, demonstrating parallels between visual perception and visual imagery. They observed that mental images resemble *percepts*, and that mental transformations of images resemble observable changes of things in the world. Later on, Johnson-Laird (1983) argued that the view of imagery as an internalized perception was too narrow and could not account for syllogistic reasoning. He proposed that people form mental models of the situations described in the representations. These models differ from classic images in that they are more schematic: entities are represented as tokens, and spatial relations are approximate. Once internalized, representations can be promptly manipulated.

People using visuospatial representations for reasoning need to manipulate their features. Transformations include moving parts of a figure, rotating them, removing or adding parts, changing size, color or shading of a component of the initial representation. These manipulations are useful to perform comparisons between parts of a representation (e.g., distance, orientation, shape, size, etc.) or to determine static or dynamic properties of entities (e.g., symmetry, texture, speed, etc.). The cognitive load associated with these manipulations is not equivalent (Kosslyn et al., 1978).

The availability of these mental representations of space or objects in space, sometimes called *cognitive maps*, allows for the making of inferences. People are quite competent in making spatial inferences, like taking a listener's point of view of a scene, when giving directions. Tversky (1998) noted how the processes underlying spatial inferences are different for the person's immediate surroundings and for larger environments. Tversky's experiments confirmed the *spatial framework theory* according to which people construct a mental spatial framework for their surroundings from extensions of three axes of the body, head/feet, front/back, and left/right (Franklin & Tversky, 1990) (see figure 2-6).

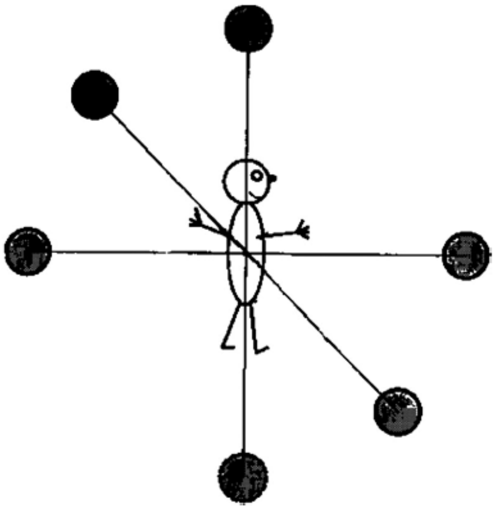


Figure 2-6: Spatial framework situation. Participants read a narrative describing objects around an observer, after Tversky, 2005

However, the focus of this thesis is on larger kinds of environments, typically geographical environments for which a cartographic representation might yield interesting cognitive applications. Over the years, research revealed many systematic distortions encoded in internal representations of maps. For instance, distances between two points in a town are often overestimated depending on the complexity and the amount of information between the two points on the map[2] (Newcombe

[2] Other systematic distortions are described by Tversky (2005).

& Liben, 1982). Recent studies revealed that is unlikely that people maintain a library of "cognitive maps" that they consult when needed. Rather, it seems that people construct representations on the fly, incorporating the minimal set of information necessary to formulate a judgment. This information comes from different sources: some may be visuospatial, acquired through experience or from maps. Other sources may be linguistic. Tversky suggested that *cognitive collage* seems to be a more apt metaphor than map for the representations that underlie spatial judgment and memory (Tversky, 1993). Such collages are schematic, they leave out much information, as it may be unknown or unnecessary, and simplify others. Schematization always entails systematic errors[3].

Schematization of reality always occurs also in the context of graphics, maps, diagrams, *et similia*, which are often the objects of spatial reasoning studies. Graphics consists of elements and spatial relations among the elements. Representations may strive for fidelity of likeliness in relation to the represented reality or can use symbolism or rhetorical strategies. Relation among entities is usually represented using space. Ordinal information, like a time series, may be represented using a list of items. Space can also be used to represent interval or ratio information.

Maps are special kinds of diagrams. They have a *scale* that is used to communicate distances, a *projection*, that is used for communicating directions and finally a set of *abstract signs*, a part of which may be text, for communicating the semantic meaning of landscape features (see an example of map in figure 2-7). The scale, projection ad and array of signs do not need to be explicitly defined or translated into written words in order to be understood. If all these signs are pictorial, or iconic, the need of a legend that acts as a dictionary for interpreting the content might become less important.[4] Also, the properties of scale and projection can be inferred, at least roughly, from the image. Blaut and Stea (1971) conducted pioneer studies on the abilities of children aged five through ten to understand maps. They found that map learning begins long before the child encounters formal geography and cartography. They demonstrated that preliterate children of five and six can deal with map-like representations. The mapping language, in its elementary forms, is independent of the written natural language. This finding seems consistent cross-cultures (Blades et al., 1998).

Uttal (2000) studied how the use of maps, from dirt drawings and stone carvings to topographic map sheets, appears to be a cultural universal. Maps record what is known and remembered about an environment and act as a support to wayfinding. In the absence of these artifacts, people rely on internal representations, or memories, of experienced environments (Golledge, 1999). Concerning the acquisition of spatial information, one of the most general characteristic of maps is that they lead their user to acquire a knowledge of the world that exceeds their direct

[3] See for instance Milgram's study on the way Parisians mentally represent their city (Milgram, 1976). It is not a representation of Paris as a geographic reality, but rather of the way that reality is mirrored in the minds of its inhabitants. And the first principle is that reality and image are imperfectly linked.

[4] However, it must be noted that even with pictorial language, the danger of miscomprehension is still present.

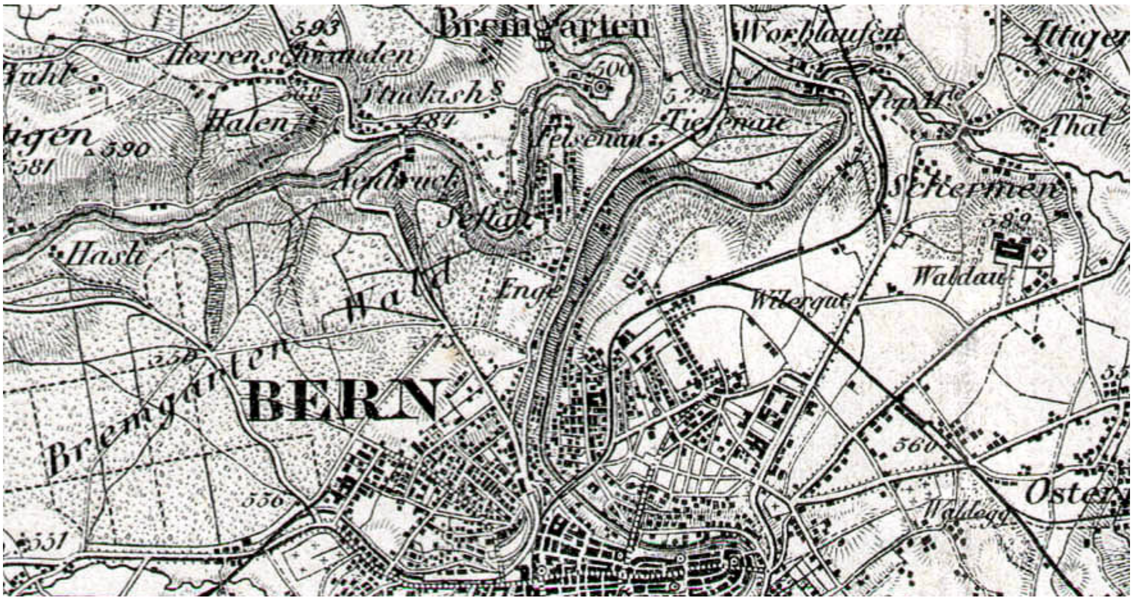


Figure 2-7: Historical map of Bern. It was realized in 1907 by Dufour. Elevation is represented with orthogonal marks rather than isobars.

experience. Maps allow for the exploration of spatial relations without the need of navigating through the space (D. Wood, 1992). Maps help to realize that it is possible to think and represent the world beyond a person's direct experience. Of course, the reality provided by maps is not free from distortions and biases. Maps also influence how we acquire spatial information. While navigating, we constantly change the relation between our position and our viewpoint. Therefore, the different spatial features of our surroundings change their perceptual salience. In contrast, maps provide a static point of view over space (E. Hutchins, 1995). Additionally, because the graphical conventions of maps are somewhat arbitrary, they can artificially change the salience of spatial features. Subsequently, maps allow users to gain visual access to a number of spatial relations that would not be available through direct experience (Uttal, 2000). In synthesis, maps give access to spatial and geographic information that would otherwise be inaccessible through actual navigation of the space. This discussion therefore suggests that maps have strong cognitive consequences in the way people think about the space. Maps give people the possibility of thinking about the space in map-like terms. People may form mental models of city-wide space that are influenced by their experience of working with maps (Liben, 1999).

Extending these findings, Uttal (2000) suggested that the relation between maps and the development of spatial cognition is reciprocal in nature. Numerous studies have shown that providing students a geographic map of a place as an adjunct to text descriptions of the same place lead to better recall of text information than when students are provided only the text descriptions (Kulhavy & Stock, 1996). These findings may support the assumption that information is stored both visually and spatially. Paivio (1990) proposed that performance in memory and other cognitive tasks is mediated not only by linguistic processes but also by a distinct nonverbal imagery model of thought.

More generally, Bauer and Johnson-Laird (1993) demonstrated that using a spatial diagram facilitates temporal problem solving. Mayer and Gallini (1990) also showed how diagrams can support many different classes of inferences, notably *functional* and *structural*. Structural inferences are inferences about qualities of parts and the relation among them (e.g., distance, direction, size). These qualities can be extracted easily using the transformations discussed above and without any additional knowledge or expertise. Conversely, making a functional inference requires linking specific perceptual information to conceptual information: this combination corresponds to the *mental model* of Johnson-Laird (1983). In this regard, Suwa and Tversky (1997) showed how architects are more at ease in making functional inferences from architectural diagrams than novices, but that this difference was not measured for structural inferences.

Paper-based diagrams are useful for supporting annotations. People like to annotate maps for a variety of purposes: remembering specific locations or paths, communicating these locations or paths to others or calculating routes or other measures for which distance or position plays a role (see figure 2-8). Of course, people also annotate maps to add information on resources that might be available at specific locations. Map annotations are powerful tools for supporting collaboration in presence or at distance, at the same time or at future events. However, there are inconsistent findings on whether enriching diagrams with extra-pictorial devices might facilitate functional inferences. Tversky et al. (2007) suggested that the lack of standards in designing and using extra-pictorial devices is detrimental to the use of diagrams and maps in general, and to the inference process that can arise from their use.

While scientific diagrams and maps are designed to communicate clearly, efficiently and without errors, sketches and artistic drawings are sometimes created to be ambiguous, to allow for discovery and reinterpretation. The process of using a sketch for some unintended discoveries is termed *constructive perception*. It consists of two independent processes: the mental reorganization of the sketch, and the process of relating the new organization to a design purpose (Tversky, 2005).

Sketches have been studied extensively in many design disciplines (Henderson, 1999). There are a number of findings that may be relevant to the study discussed in this thesis. For example, it has been observed that designers and engineers sketch for four different but related reasons:

- *To share*: Diagrams play a major role in communication (Tversky, 2001), as they externalize internal models through making it visible to self and others (Tversky et al., 2003), reifying the mental model for others to act upon.

- *To ground*: Human communication embeds ambiguous interpretations that need to be clari-

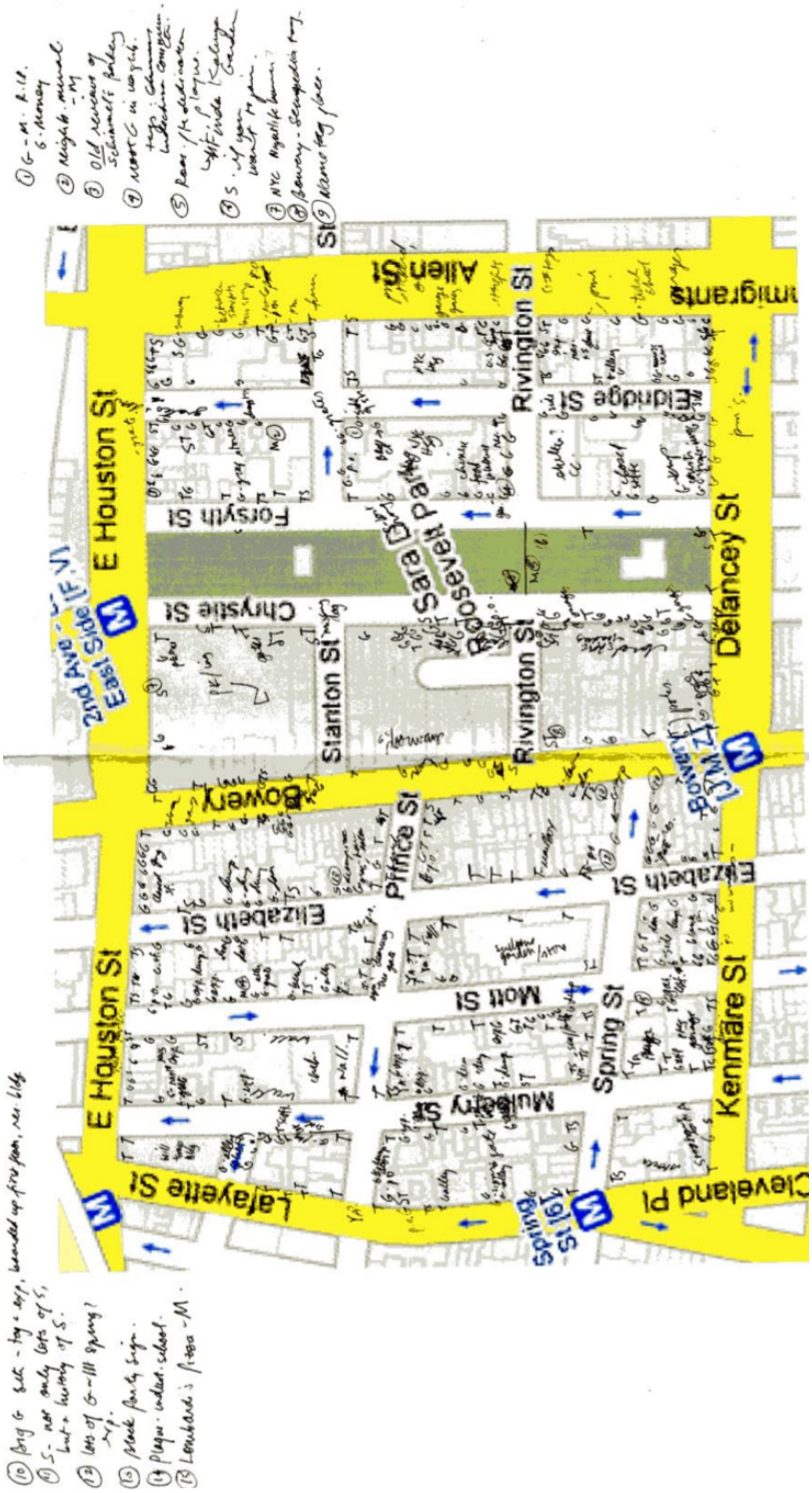


Figure 2-8: Example annotate survey map (from Kottamasu, 2007, p. 41)

fied in conversations (Cherubini et al., 2005): diagrams can become a means of clarification.

- *To manipulate*: By externalizing a mental model in a drawing, part of the cognitive process needed to hold it on memory is relieved and other operations can take place, like joining different parts, evaluating the design, checking the consistency, etc. Once externalized, these phases can happen collaboratively, capturing joint attention and enabling gesturing (Alibali et al., 1999; Goldin-Meadow, 2003).
- *To brainstorm*: Ambiguity in sketches is a source of creativity. Unintended interpretations and ideas can arise when inspecting an initial arrangement of a sketch (Suwa et al., 2000).

The cognitive implications are manifold: diagrams support communicating, capturing attention and grounding conversations (Clark & Shaefer, 1989). They reduce the cognitive burden of evaluating a design or considering new ideas (E. Hutchins, 1995).

2.2.3 Synthesis: space descriptions and representations

People use a limited number of different perspectives when they need to describe space. Depending on the scale of the environments they want to talk about, they employ different strategies. A small environment like a doll-house is usually described with a gaze tour, guiding literally the attention of the listener through the different elements to be described. If the speaker needs to describe the space-at-sight around the body, then a self-referred coordinate system is usually employed. Finally, when the environment is *"too large to be seen in a glance"*[5], and depending on the number of features to be encoded, a different strategy like a route or a survey perspective is usually taken. People also are flexible in their chosen strategy. They often mix perspectives in their descriptions. Also, the choice of what strategy to use is largely dependent on the task at hand and on individual and linguistic styles.

The spatial information necessary for interaction is schematized in the mind of the interactants. A schematization always entails reduction. A cognitive collage is assembled from the spatial scene, in which irrelevant or unhelpful information is discarded. Often this simplification of information introduces systematic errors. Schematization and reduction also always occur with externalized imagery such as diagrams and maps.

Several studies have shown the usefulness of diagrams in comprehension and problem solving. Many more described how sketches are indeed helpful in design because they externalize mental representations, and they allow for manipulating, sharing, and brainstorming ideas. However, few studies investigated how adding extra-pictorial information to diagrams, and maps in particular, could be used in collaborative situations and what the corresponding cognitive benefits would be.

[5] This expression was often used by Tversky to describe environments such as cities, campuses, and so forth.

2.3 How physical space influences cognition and interaction

Although *space* is a continuous realm, human action in space is conceptualized as being discrete and bounded. This is indeed the argument of Harrison and Dourish (1996), who advocate talking about *place* rather than space. People act in places. While space can be defined with quantitative parameters such as coordinates or extensions or distances, places are defined by qualitative properties. The person inhabiting places labels them "home", "work", "gym", etc. The function of that space for its 'owner' determines these names and they evolve over time. Places should be examined in terms of the social matrix that gives them meaning. Space becomes place through the incorporation of social actions, norms and a cultural understanding of use (Nova, 2005). Ten years after, Dourish reconsidered and adjusted his initial arguments, suggesting that space is as much social product as place is: the conceptual resources available when talking about space are the products of particular kinds of social practice (Dourish, 2006, p. 301):

> I have argued that the predominant interpretation of the relationship between place and space has looked at space as pre-given and place as a social product. From that point of view, the overriding technical question is to understand those features of spaces that are conducive to the creation or emergence of place. However, I have argued for a different perspective, one that recognizes the ways that both space and place are products of embodied social practice.
>
> What this suggests, then, is that I need to understand, first, something of the relationship between spatiality and practice, and, second, how multiple spatialities might intersect. This is particularly the case when I think not about "virtual" settings but rather about the ways in which wireless and other technologies might cause people to re-encounter everyday space. Introducing technology into these settings does not simply create new opportunities for sociality (the creation of places); rather, it transforms the opportunities for understanding the structure of those settings (developing spatialities).

This is of particular interest for the collaborative annotations application that is the core of this thesis because annotations are, in essence, defined by spatial features (e.g., latitudes and longitudes), but they define, or relate to, places rather than spaces. Therefore, studying spatial annotations yields an understanding of how people make sense of a specific location.

2.3.1 Partitioning space

Humans also partition virtual spaces to define a particular domain of interaction. A virtual brainstorming room can relate to work, while a virtual room named "John's pub" can relate to leisure, for example (Benford et al., 1993).

Knowledge of partitions and the presence of our collaborators in particular places will prompt a number of inferences that will be used during the collaboration process. For instance, if a person is known to be in a personal space, she might be interested in discussing activities for the weekend, while if the same person is known to be in a meeting area, then s/he might not be available for such a conversation.

Researchers were also interested in the person to person relationship in space and how these affected collaboration. Proximity has proved to improve communication processes. Communication and therefore collaboration is easier in physical settings than through computer mediated contexts. When co-located collaborators meet more frequently, they are more likely to feel part of the same community and have a better established group awareness (i.e., who is doing what). Many meetings opportunities are triggered by repeated or serendipitous encounters (Kraut, Fussell, et al., 2002).

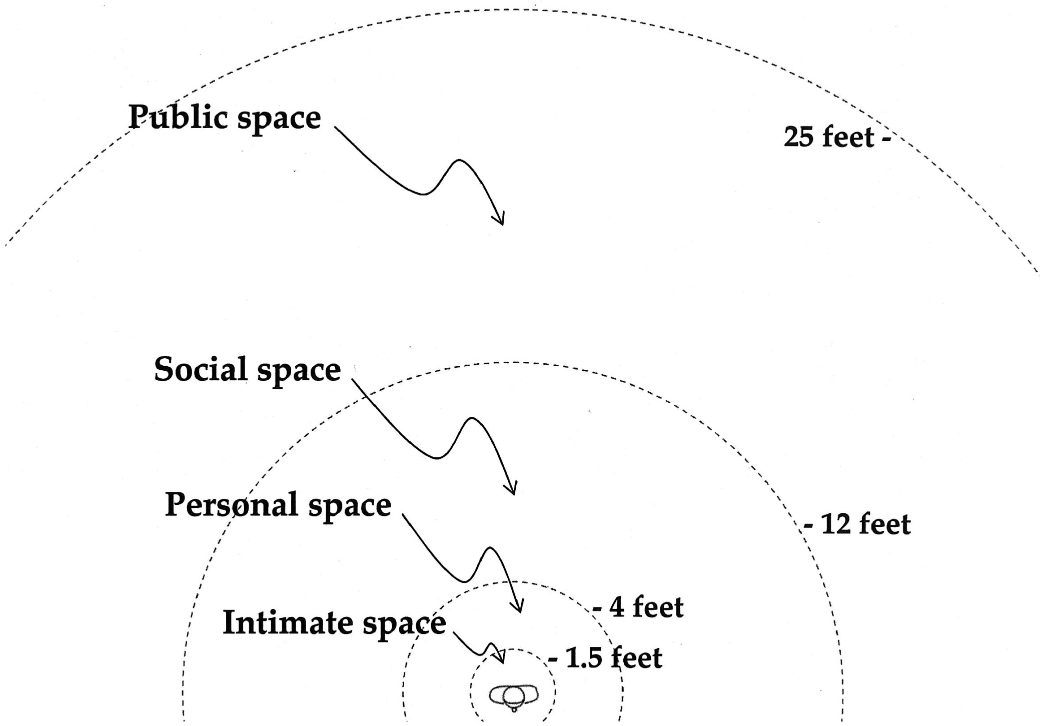


Figure 2-9: Diagram of Hall's personal reaction bubbles, showing radius in feet, from E. T. Hall, 1966

Hall additionally studied how space structures social interaction in what he defined *proxemics*: the distance between people is a marker that expresses the social relation between the parties (E. T. Hall, 1966). Hall proposed four different spheres that afford distinct types of interactions (see figure 2-9). A *public distance*, used for public speaking, ranges from 3.6 to 7.5 meters (12 to 25 feet); the *social distance*, used for interactions among acquaintances, ranges from 1.5 to 3.6

meters (5 to 12 feet); a *personal distance*, used for interactions among good friends, ranges from 45 centimeters to 1.2 meters (1.5 to 4 feet); finally an i*ntimate distance*, used for embracing, touching or whispering, ranges from 15 centimeters to 45 centimeters (6 to 18 inches). The classification of a relation to a certain sphere was based on the actual distance between the bodies of the interactants. He also demonstrated how these distances were culturally dependent and that this body distance was truly related to the occurring interactions. Therefore, the position of people in space communicates a lot about the nature of the relations between the participants and their activities to the interactants as well as to observers.

2.3.2 Artifacts and interaction

What is of more interest to this thesis, though, is *the relationship between person and artifacts* in the interaction space. In co-located meetings, it is well known how conversation participants take advantage of the artifacts located in the vicinity of where the conversation takes place to avoid miscomprehension. Krauss and Weinheimer named this behaviour as *referential communication* (Krauss & Weinheimer, 1966). Pointing, looking and gesturing are an essential part of human communication. When this is done to indicate a specific object, or a point of interest in the surroundings, then it is called deictic reference. Little research has addressed the use of referential communication for collaboration in virtual space, as I will discuss more in details in chapter 3.

The spatial environment is a resource in communication and it is also a resource in collaboration and in problem solving. According to Kirsh and Maglio (1994; 1995), actions like pointing, annotating, manipulating artifacts and arranging the position and orientation of nearby objects are examples of how people encode the state of a process or simplify perception.

One last note concerns the importance of *territories*. Human territoriality reflects on the personalization of an area to communicate ownership. Prohansky et al. (Proshansky et al., 1970) found a strong relation between group identity, namely the feeling that we belong to the same human group, and spatial identity, namely the experience and knowledge of the surrounding environment. The authors also observed how territoriality is a way to achieve and exert control over a segment of space and then to achieve and maintain a certain level of privacy. Jeffrey and Mark (1998) observed how the same human behavior is expressed in virtual worlds, where participants who can potentially dispose of an infinite amount of space tend to build their virtual-houses in virtual-neighborhoods.

2.3.3 Spatial awareness and collaborative work

Location awareness is the knowledge of the position of one's interaction partners both in physical environments and in virtual worlds (Dyck & Gutwin, 2002). This information is extremely important for the coordination of communication and collaborative problem solving especially when participants are not nearby. Nova conducted a series of timely experiments to demonstrate the impact of location awareness on collaborative work at a distance (Nova, 2007). He designed an ubiquitous treasure-hunt game, called CatchBob!, where three participants had to walk around a campus area and chase a virtual object. They had at their disposal a tablet PC running the interface represented in figure 2-10. Using this system, they could communicate with their partners annotating the campus map with the stylus of the tablet and have information on their proximity to the virtual object to be found. He compared experimental conditions where partners could see the position of their partners with a control condition where participants could see only their position. After the game, he interviewed participants, and asked them to draw the recalled path of his/her partners, comparing this information with the real traces from the system logs.

Using this experimental design, he demonstrated that the availability of what he called Mutual-Location Awareness tool (MLA) had an impact on collaboration. In particular, the knowledge of where the other players were located had inhibiting effects on communication within groups and on the recall of partners' past positions both with automatic and manual refresh of the information. It also made the group more passive than those who did not have this interface. Nova and colleagues also analyzed the messages exchanged during the game. He used the coding scheme reported in figure 2-11, where messages were classed by content and by pragmatic use[6]. Consistently with the process measures described above, he found that players in the control condition exchanged more messages about position, direction and strategy that those with the MLA tool. Nova and colleagues also found a negative correlation between the frequency of messages about strategy and the number of errors made by the individual when drawing their partner's path (Nova et al., 2005).

Gutwin and Greenberg (1999) investigated the use of Mutual-Location awareness in groupware. The authors tested how the presence of workspace miniatures in the form of a map plus a telepointer (see next chapter) would affect the group task performance. They found that completion time was lower with the MLA interface in two tasks, and in a third one, communication was less efficient.

These studies highlighted the various roles of mutual location-awareness ranging from a resource for division of labor to the facilitation of situation understanding or the use of past positions to draw hypotheses about the partners' future behavior. Nova finally discussed how automating location-awareness could be detrimental to group collaboration in certain situations (Nova, 2007).

[6] The interface of CatchBob! supports Explicit Referencing, as defined in this thesis. It allows its user to write a message over the map right on the place where this information might be needed.

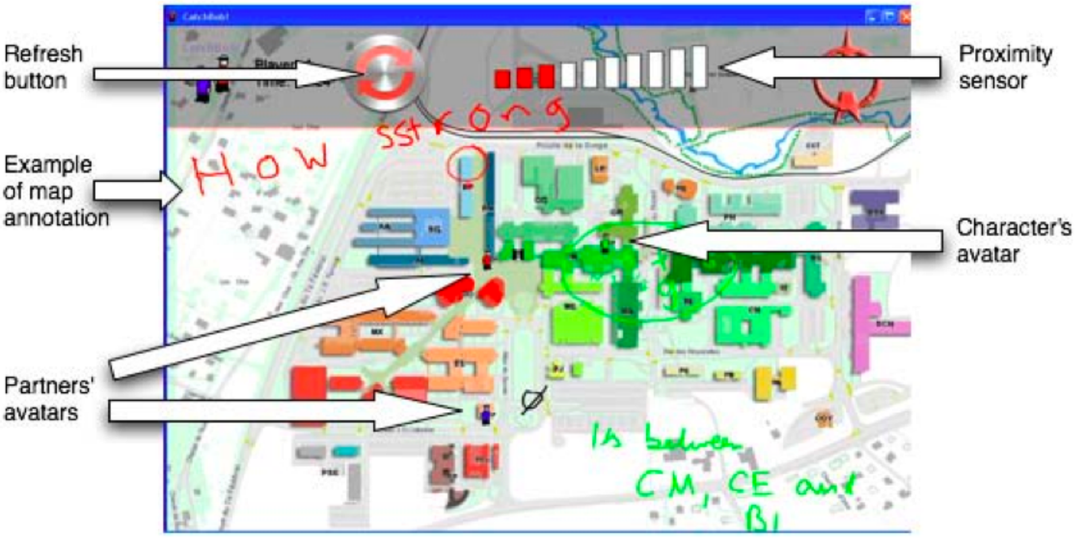


Figure 2-10: CatchBob! interface as seen by one player (from Nova, 2007)

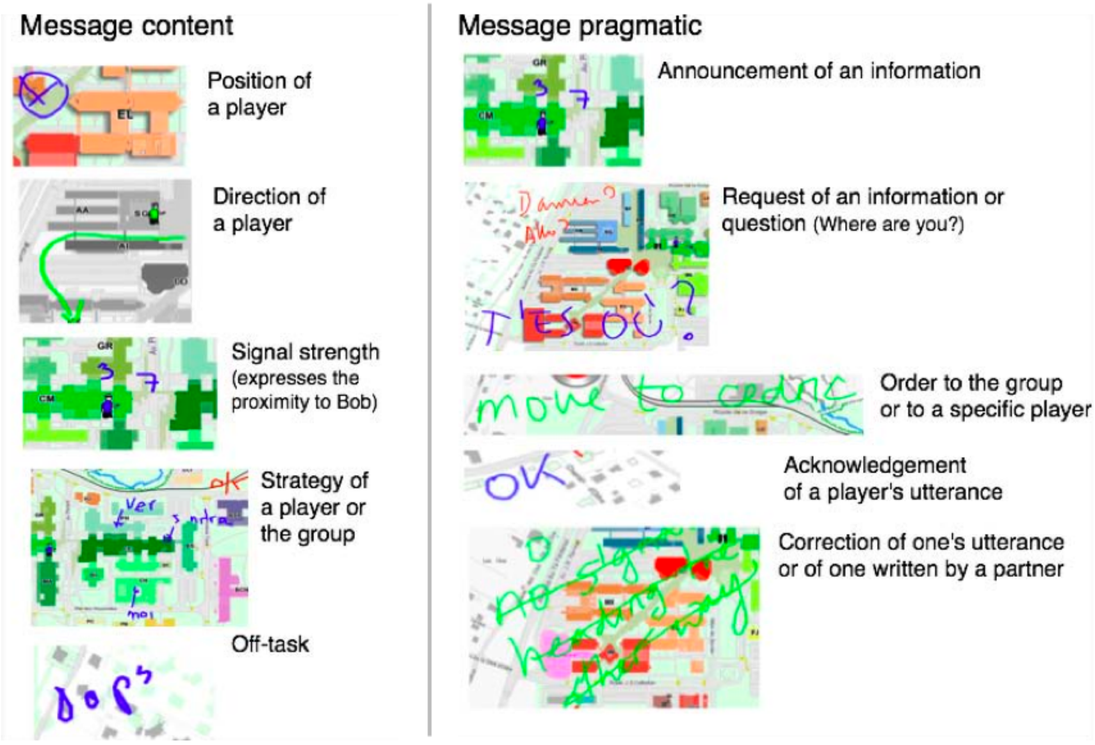

Message content
Position of a player
Direction of a player
Signal strength (expresses the proximity to Bob)
Strategy of a player or the group
Off-task
Message pragmatic
Announcement of an information
Request of an information or question (Where are you?)
T'ES OÙ ?
Order to the group or to a specific player
move to cedric
Acknowledgement of a player's utterance
OK
Correction of one's utterance or of one written by a partner
no stop heading the way

Figure 2-11: Example of messages exchanged by the CatchBob! players. Nova coded these messages using two intertwined coding scheme: message content and message pragmatics (from Nova, 2007)

2.3.4 Space and mobile communications

Communication technologies allow users to interact at a distance and from changing locations. What is relevant for me here is that this possibility undermines interaction as the context in which the conversation happens when speakers are in the same place at the same time is fractured into different situations when conversation participants are separated in space, or worse in time. Therefore, mobile conversants need to reconstruct the broken context of their conversation. One strategy that is used to this end is the giving of a geographical formulation as part of an opening of a phone call. Laurier (2001) highlighted how these "location formulations" allow dispersed cell phone users to mutually establish and share a spatio-temporal context.

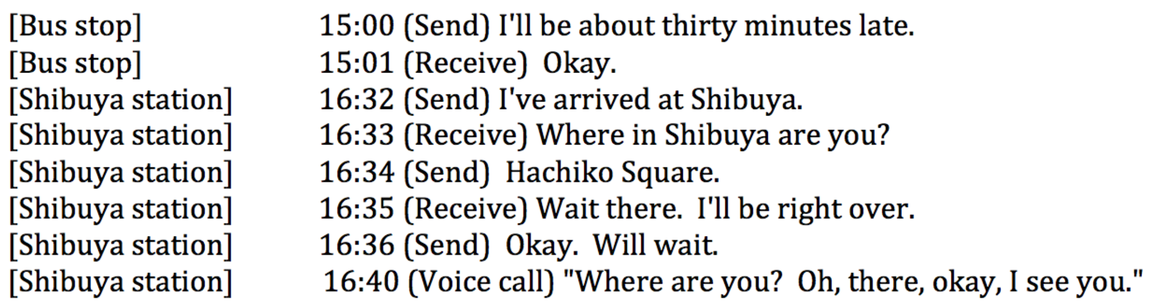
Arminen (2006) conducted several ethnographical observations of mobile phone conversations, describing three "social functions" of location. According to this author, location is used to evaluate the availability of the answerer, to negotiate a practical arrangement, and finally to share socio-emotional content. Inferences drawn from location are not very often discussed but they are used to negotiate further actions and coordination issues. Additionally, Cooper (2002) found that providing one's location seems to orientate the content of the message as well as privacy issues (p. 26):

> ... information on the whereabouts often serves to establish the grounds for the conversation in terms of constraints and sensitivities with regard to possible topic, privacy duration and so forth.

Colbert (2007) conducted a diary study of university students' use of mobile telephones for rendezvousing (arranging, traveling to, informal meetings with friends and family). His study reveals, and suggests explanation for the number of deficits in user performance, particularly that demonstrates rendezvousing outcomes are worse when meeting at unfamiliar locations. When mobile phones are used on the move, the experience of communication is slightly worse than when phones are used prior to departure. Finally, he suggested a number of solutions to overcome these deficits. He proposed an automatic and controlled disclosure of position to respond to the occasional high stress recorded by the participants of his study. Also, he suggested a location-based reminder system to help the elderly to cope with en route activities.

More interest to this thesis is how people use location to disambiguate SMS[7] content. Mobile textual messages are a mass phenomenon. Youths use them every day for chatting, coordinating and planning activities (Grinter & Eldridge, 2003). Different research tried to shed light on how this form of communication is used (Kasesniemi & Rautiainen, 2002; Ling, 2001). Some interesting findings report a major problem related with texting: the difficulty of disambiguating

[7] Short Message Service (SMS) is a communications protocol allowing the interchange of short messages between mobile telephony devices.



[Bus stop]	15:00 (Send) I'll be about thirty minutes late.
[Bus stop]	15:01 (Receive) Okay.
[Shibuya station]	16:32 (Send) I've arrived at Shibuya.
[Shibuya station]	16:33 (Receive) Where in Shibuya are you?
[Shibuya station]	16:34 (Send) Hachiko Square.
[Shibuya station]	16:35 (Receive) Wait there. I'll be right over.
[Shibuya station]	16:36 (Send) Okay. Will wait.
[Shibuya station]	16:40 (Voice call) "Where are you? Oh, there, okay, I see you."

Figure 2-12: An example of SMS exchange used for coordinating a meeting (the excerpt was taken from Ito & Okabe, 2005, p. 11)

the utterances of the messages with the available contextual information, which is usually scarce (Grinter & Eldridge, 2001).

Ito (2005) in particular, analyzed the use of mobile phones in Japanese youths. Instead of considering mobiles as a technology that disrupt current social practices, the author propose a view of phones that create new kind of boundless places that merge the infrastructures of geography and technology, and new kinds of technological practices that merge technical standards and social norms. They call these "technological situations", a way of incorporating the insight of theories of practice and social interaction into a framework that takes into account mediated social orders. They presented three different technological situations that are built on mobile email: *1. mobile chat*, an analogous of text-chat used to fill dead time; *2. ambient virtual co-presence*, a way of maintaining background awareness of others; and the *3. augmented flesh meet*, a way to augment the experience of physically co-located encounters. People exchanging messages in the second scenario used SMS as a secondary channel of communication, to maintain an awareness of what their peers are doing and where they are. SMSs have the advantage that they can be used in context where voice conversations might be inappropriate. People in the third scenario exchange SMS to enhance their face-to-face experience: to bring in the presence of the other who did not make it to the physical gathering, or to access information that is relevant to their meeting place or time. In these observations, location information was provided if the sender had a doubt that the message could be misunderstood. Also, as messages have a limited length, providing the location of emission of the message or related to future positions of the sender could prompt the receiver to make inferences about the context in which the message was written and therefore reduce the number of words necessary for the sender to make his/her point. Little research has explained how people use spatial expressions in SMS. Some interesting findings come from ethnographical studies like those conducted by Grinter and Eldridge (2001) and Ito (2005). For instance, one of the simplest scenarios observed by Ito was that of using SMS to coordinate for a meeting (see figure 2-12).

2.3.5 Synthesis: space, communication and interaction at a distance

– Space and places are a fundamental resource that people use to structure interaction and to propel collaboration. The knowledge or the provision of one's position can be used to make interaction more effective. In particular, when participants are communicating at a distance, this information can be extremely valuable in reconstructing the broken context of interaction.

– Automating the provision of location information, creating a sort of spatial awareness for group work, was shown to have effects, both positive and negative, on collaboration. The introduction of this extra tool for supporting group work modifies the resources available to the collaborators and therefore the way they interact.

– Few works have explained the role of space in textual mobile communication. From ethnographical observations, I know that textual messages are combined with the provision of location information to support informal group activities. However, more research is needed to understand how text and space can support collaboration.

Concerning this last point, the recent development of Location Based Services gave rise to a plethora of practices of sharing geo-localized messages (virtual notes pointing to physical locations) or other forms of location-based content (Morgan, 2005; Meyer, 2004). Socialight[8], for instance, allows users to share urban landmarks by posting a picture of the place they want to share plus some textual description with their mobile phone. This information is then made available on the Internet and on the mobile phones of the user's friends. In chapter 4, I will review tools and interfaces that allow this kind of communication which combines text and spatial information.

2.4 Grounding theory between utterance and knowledge level

Many studies conducted in different disciplines highlighted the importance of referencing in collaborative work. Ethnographical studies of how people use language in relation to space, reported in the previous sections, demonstrated how this interplay is embedded in every language and how humans take advantage of the context to sharpen communication. Studies developed in the context of Human-Computer Interaction, which will be described in chapter 3, valued the importance of gestures and implemented solutions to enable remote deixis. Evaluation of such devices was rated positively. In chapter 4, I will describe more specific solutions developed in contexts ranging from urban studies to art and design for joining communication contributions to elements of a shared workspace that were also rated positively in formal evaluations. These

[8] http://www.socialight.com, last retrieved January 2008.

studies developed under various theoretical frameworks explain or support the importance of referencing mechanism. However, each theory explores the role of referencing from its own perspective and employing its own methodology. A methodological choice for the work presented in this thesis is therefore needed.

My concern is to investigate how Explicit Referencing, as defined in the introduction, influences collaborative work at a distance. Coordination is one of the basic mechanisms of collaboration and it is achieved through communication. Therefore, I need to address human-human communication choosing a framework that will give both conceptual constructs and explanatory mechanisms to describe how explicit referencing impact coordination. For this reason, I chose a psycholinguistic framework, and particularly, I adapted a research framework that was defined by Nova in his investigation of the role of location awareness in Computer Supported Collaborative Work (Nova, 2007), which was in turn derived directly from Clark's theory (1996). This framework consider verbal and non-verbal communication as actions and therefore as observable aspects of cognitive processes.

Next, I will present the elements of this framework. Of course, Clark's theory received some criticism over the years that will be addressed in section 2.4.2. Particularly, the section will show how the application of this theory to CSCW poses some problems and a possible adaptation fo Clark's theory will be proposed in section 2.4.3.

2.4.1 Research framework derived from Clark's theory

During the last thirty years, Clark developed a theory of how people coordinate their actions using language (1996). The theory that he developed with his colleagues was often referred to as the *grounding theory* (Clark & Shaefer, 1989; Clark & Brennan, 1991). Clark argued that, when a group of people have to achieve a *common goal*, they need to perform *joint activities* that require coordination among the group's members. Joint activities are composed of set of identifiable tasks that Clark named *joint actions* or *participatory actions*. These actions are "joint" because each person's action is dependent on the actions of the other. People acting jointly always face *coordination problems* (Schelling, 1960).

To solve coordination problems, people need to share *coordination devices*, a rationale for mutual expectations that make partners believe that they will converge on the same joint action (Lewis, 1969). Clark expanded this definition by specifying four different kinds of coordination devices: conventional procedures, explicit agreement, precedent and manifest elements.

- *Conventional procedures* are a community's solutions to recurrent coordination problems. These range from rules and regulations to less formal codes of conduct like habits or practices (e.g., stopping when a pedestrian is crossing the street).

- *Explicit agreements* are dialogues in which parties explicitly communicate their own intentions (e.g., “I am coming with you”).

- *Precedent* coordination devices refer to norms and expectations developed within the ongoing experience of the joint activity. For instance, if two people are discussing a meeting place and one remembers from a previous conversation that the partner knows a place called “roxy pub”, then she might decide to use again this place for this meeting.

- Finally, *manifest elements* are communication devices derived from the environment in which the communication takes place. Clark defined the manifest elements as *perceptual salience*: situations in which the environment or the available information makes the next move apparent among the many moves that could conceivably be chosen. For example, if an utterance is ambiguous in a certain context (e.g., “take the red book on the table”, with multiple reddish books on the same table), this can be easily disambiguated by joining a deictic gesture to the contribution (e.g., “take that book”).

While conventions and precedent devices refer to mental representations, manifest elements refer to perceptual elements from the environment and explicit agreement about specific elements of the communication.

Clark (1996) defined *the principle of joint salience* as the idea that the best coordination device for any problem is that is most salient, prominent or conspicuous with respect to the *common ground* of all participants. By common ground, CG in short, Clark specifically meant the knowledge, beliefs and suppositions participants share about an activity, and that accumulate through out a course of action. Simply stated, the common ground is the accumulation of information exchanged as coordination devices. The process of constructing and updating the information in the common ground is defined as *grounding* (Clark & Shaefer, 1989; Clark & Brennan, 1991). Clark described three components of CG:

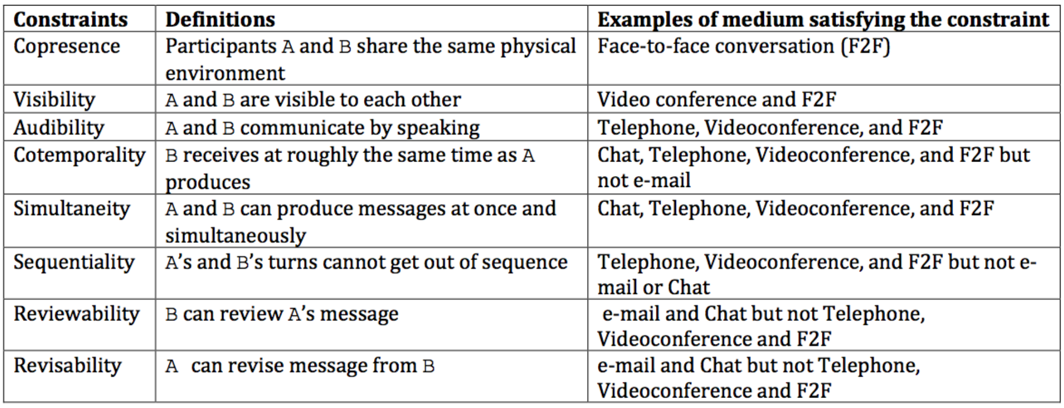
- *The initial common ground* is the set of background facts, assumptions and beliefs that participants presupposed when they entered the joint activity. Precedent and conventions belong to this category. This form of CG represents the initial context of an activity.

- The *current state of joint activity* is the participants’ up-to-the-moment understanding of the state of an activity. The environment and artifacts play an important role as external representations of the current state. This category includes the manifest elements and conventions set during the activity.

- *Public events so far* are the history of the public events incurred in the joint activities prior to the current state. Manifest elements from the situation and explicit agreements belong to this last category.

Grounding a conversation requires a collaborative effort. Clark and Brennan (1991) explained how people tend to minimize this effort to a criterion sufficient for the current purpose. This was called the *least collaborative effort* principle. Additionally, they provided a concrete account for how different communication media can impact the establishment of the common ground. They explained how coordination devices are conveyed differently in different media. For instance, while the expression "hmm hmm", is very often used and recognized as an acknowledgment in face-to-face conversation, in a chat conversation it can be easily interpreted as an interruption. Consequently, each medium has different costs, for instance acknowledging requires more effort in a chat conversation that when people are face-to-face. Table 2.2 summarizes the constraints imposed on communication by diverse media as explained by Clark and Brennan (1991). This table gives ideas on how different media can affect the cost of production of an utterance, or how difficult it might be to repair a message produced in different systems. Each constraint has positive and negative aspects. For instance, co-located interactions maximise the possibility of using gestures but chat dialogues have the advantage of providing a permanent record of the conversation.

Table 2.2: Constraints on grounding and examples (after Clark and Brennan 1991)



Constraints	**Definitions**	**Examples of medium satisfying the constraint**
Copresence	Participants A and B share the same physical environment	Face-to-face conversation (F2F)
Visibility	A and B are visible to each other	Video conference and F2F
Audibility	A and B communicate by speaking	Telephone, Videoconference, and F2F
Cotemporality	B receives at roughly the same time as A produces	Chat, Telephone, Videoconference, and F2F but not e-mail
Simultaneity	A and B can produce messages at once and simultaneously	Chat, Telephone, Videoconference, and F2F
Sequentiality	A's and B's turns cannot get out of sequence	Telephone, Videoconference, and F2F but not e-mail or Chat
Reviewability	B can review A's message	e-mail and Chat but not Telephone, Videoconference and F2F
Revisability	A can revise message from B	e-mail and Chat but not Telephone, Videoconference and F2F

As the focus of this work is on referencing mechanisms that are used to support communication, I will develop one particular aspect of Nova's original framework, namely the definition of the manifest coordination devices with respect to Clark's theory. In recent work, Clark (2003) explains how communication is ordinarily anchored to the material world and that one way it gets anchored is through *pointing*. Clark also explains that the counterpart of pointing is *placing*. Through the use of our position and the position of the objects I refer to in the actual world, I shape context to reduce misunderstanding and make communication more efficient. He argued that *directing-to* and *placing-for* are two *indicative acts*.

Indicating has fundamentally to do with creating indexes for things. Clark explains how every indication must establish an intrinsic connection between the signal and its object. The more transparent this connection is, the more effective is the act. Indicating an object in space leads the participants to focus attention on that object, or in other words, anything which focuses the attention is an *index*. Finally, every indication must establish a particular interpretation of its object. That is why an indication cannot stand on its own, independently of an associated communicative intent. Finally, I often find pointing-to and placing-for devices combined. In the same work, Clark defines what he calls a *perceptually conspicuous site*, or PCS, a site, in the shared workspace that is perceptually conspicuous relative to the speaker and interlocutor's current common ground. Gesturing often points to PCSs but this indication should always be combined with clues for interpretation, for instance, an utterance explaining the relation of the gesture with the current activity. Finally, indicating tends to be a transitory signal, while placing a continuing one.

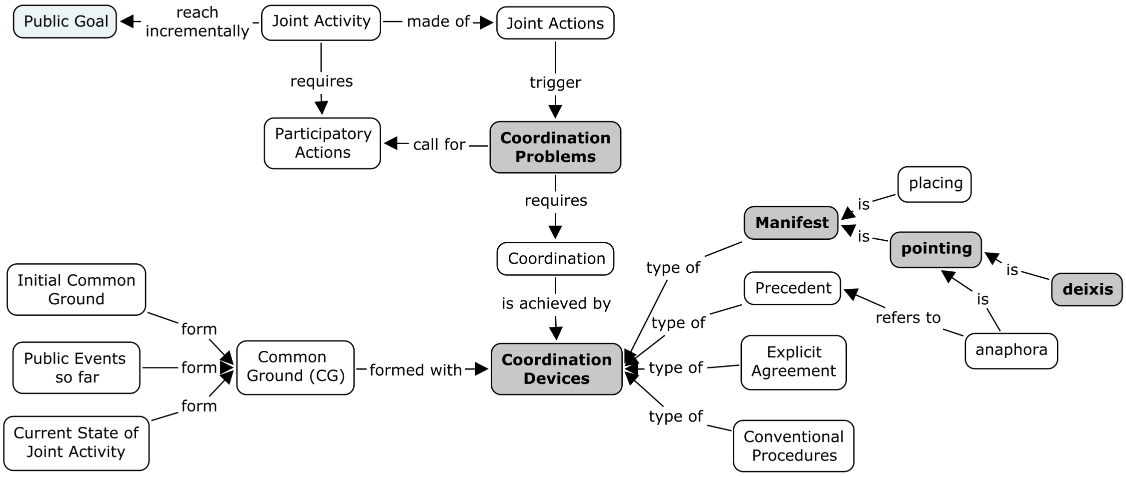


Figure 2-13: An overview of Clark's framework of coordination (adapted from Nova 2007)

One particular kind of indication is *anaphora*, a set of phenomena in which linguistic elements are missing or have been replaced by other elements (Clark & Murphy, 1982). The best known type of anaphora is pronominalization, in which a pronoun (like he) is used instead of a complete description of the referent (e.g., the tall man standing close to the door). Anaphora is a kind of indication that uses 'precedent' coordination devices.

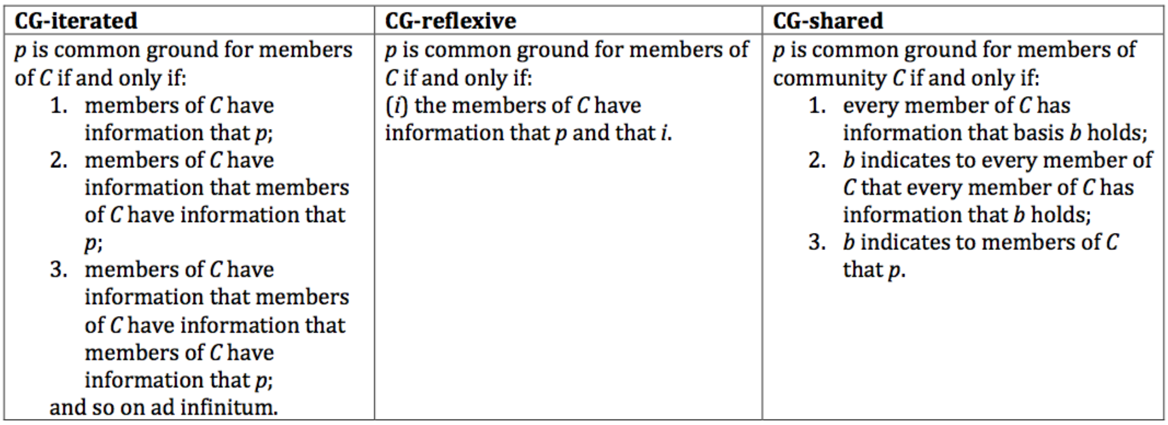
Figure 2-13 summarizes the framework I described in this section. The concepts that I have introduced (and typed in *italics*) are represented in this conceptual map, which shows the process of accumulation of coordination devices needed to solve the collaboration goal. In conclusion, while Clark's theory is widely adopted in CSCW to design or predict interpretations, criticisms have arisen over the years on the *mutual knowledge issue* by linguists of the pragmatic school. The central point of this controversial discussion is to understand exactly what is mentally shared between the conversation participants. I discuss these critiques in the next section. Particularly of interest to this thesis is the applicability of this framework to topics outside the scope of research of psycholinguistics. I focus on this issue in section 2.4.3.

2.4.2 Criticisms to Clark's theory

Clark's theory received many criticisms during the years. Smith (1982), reports an overview of these early controversies. Essentially, other scholars disagree about the following aspects of Clark's common ground: its representation (1), its reason-to-exist (2), and finally its degree of consciousness in the conversants' minds (3).

(1) Clark proposed *alternative types of representation of the CG* over the years. Table 2.3 presents three different versions that are not equivalent: **CG-iterated**, **CG-shared**, and **CG-reflexive**. In CG-iterated representation, conversants perform an infinite series of checks in order to ground certain information. Conversely, the CG-reflexive representation states that conversants have simply a meta-knowledge of their knowledge of the information (e.g., team members know that is going to rain and they know that they are aware that is going to rain). Finally, the CG-shared breaks the repeated sequence of checks after the initial iteration, simply recognizing that the knowledge of the situation and of the reciprocal awareness of the situation is sufficient to establish common ground. The CG-iterated paradigm was considered flawed in its essence as it would require an infinite cognitive capacity (Green, 1987; Schiffer, 1972). However, Clark himself declared in many papers that the CG-iterated model was not applicable to real conversations (see for instance Clark & Marshall, 1981).

Table 2.3: Three representations of common ground for a proposition *p* in a community *C* (after Clark 1996)



CG-iterated	**CG-reflexive**	**CG-shared**
p is common ground for members of *C* if and only if: 1. members of *C* have information that *p*; 2. members of *C* have information that members of *C* have information that *p*; 3. members of *C* have information that members of *C* have information that members of *C* have information that *p*; and so on ad infinitum.	*p* is common ground for members of *C* if and only if: (*i*) the members of *C* have information that *p* and that *i*.	*p* is common ground for members of community *C* if and only if: 1. every member of *C* has information that basis *b* holds; 2. *b* indicates to every member of *C* that every member of *C* has information that *b* holds; 3. *b* indicates to members of *C* that *p*.

(2) Other authors considered the notion of common ground essentially flawed. Sperber and Wilson (1986) argued that "*to achieve successful communication, CG is not necessary*". Instead, they introduced the notion of *cognitive environments* and *mutual manifestness* of facts in the participants' environment. According to Sperber and Wilson, a fact or assumption is manifest if it is perceptible or inferable (i.e., not only what a conversant is aware of but what she is capable of becoming aware of): "*A fact is manifest to an individual at a given time if and only if she is capable at that time of representing it mentally and accepting its representation as true or probably true*" (Sperber & Wilson, 1986, p. 39). For example, if two individuals have the same perceptual and cognitive abilities, a relevant fact would become *mutually manifest* to them. In this case, Sperber and Wilson state that these two conversants share the same *cognitive environment*. The central claim of their relevance theory is that the expectations of relevance raised by an utterance are precise enough, and predictable enough, to guide the hearer towards the speaker's meaning. They explain what these expectations of relevance amount to, and how they might contribute to an empirically plausible quantum of comprehension in cognitively realistic terms. The difference with Clark lies in the fact that stating that two persons share the same cognitive environment does not imply that they make the same assumptions, but merely that they are capable of doing so.

(3) Finally, other scholars countered the conception of establishment of CG as active and intentional. Pickering and Garrod (2006) argued that alignment is the basis for successful communication in dialogue. They proposed an account of the mechanisms that interlocutors employ during dialogue, according to which they *align their linguistic representation* during the interchange, with successful communication occurring when they become well aligned. They developed a theory of alignment where automatic processes play a central role and explicit modeling of one's interlocutor is secondary (Pickering & Garrod, 2004). Alignment is generated by primitive priming mechanisms that requires no processing effort and entails no explicit negotiation between interlocutors. According to this theory, interlocutors do not model each others' mental states. Instead, they simply align on each other's linguistic representations. Additionally, Pickering and Garrod (2006) proposed that successful dialogue involves the alignment of situation models through three processes: (a) the automatic mechanisms of linguistic alignment priming on linguistic representations; (b) alignment repair mechanism; (c) alignment via explicit modeling, which is used as last resource. While Clark accepts that interlocutors tend to converge on the same expressions, he argues that this convergence is not necessary for the construction of common ground. Conversely, Pickering and Garrod accept that in principle communication can be successful without linguistic convergence, however they argue that it would be extremely impoverished in practice.

While the first two criticisms of Clark's theory can be countered stating that they focused on earlier versions of his framework, I cannot say the same for the last criticism which only recently appeared in the literature. Additionally, I can say that the notion of coordination devices exchanged during an interaction is not far from the notion of mutual cognitive environment. More precisely, if we consider only the elements of the common ground that are derived from the environment, both manifest and perceptible elements, this local CG would correspond to the mutual cognitive environment. As Nova suggested (2007), we can consider Clark's coordination devices as *relevant elements for inferences*, using Sperber and Wilson's terms. People's communicative coordination in a collaborative effort will depend on the choice of the most relevant interpretation among possible interpretations of each other's contribution. Participants operate this choice, considering the relevance of the coordination devices exchanged. Finally, these debates are still ongoing and their resolution is beyond the scope of this discussion. I reported them here to offer alternative views on the framework that will be used in this thesis. Next, I will discuss my own criticisms of Clark's theory, which are more relevant for this work.

2.4.3 Four dimensions of grounding at a micro and macro levels

(*)[9] Many studies of Computer Supported Cooperative Work that identify grounding as an important process, analyse it using the theory of (or models based on) Clark and Shaefer (1989). However, the application of their theory within the field of CSCL and CSCW has some problems. As a linguistic theory, it analyses conversation on a micro or 'utterance' level and is not developed to describe the macro or 'knowledge' level, which is associated with collaborative work. While the micro level focuses on the dialogue interchange occurring between two or more interlocutors, usually of a short length (e.g., 2 seconds), the macro level refers to the shared understanding that is constructed as a consequence of that exchange (Dillenbourg & Traum, 2006), typically over a longer sequence (e.g., 20 minutes). I argue that the observable presentation and acceptation of utterances, as described in Clark and Shaefer's contribution theory, cannot automatically be translated into the sharing of knowledge. As Koschman's (1996) example of a learning conversation between surgeon and student in an operation room shows, even repeated presentation and acceptation phases of a concrete referent in a shared environment, can result in different personal representations at a knowledge level. More recently, Koschman and LeBaron (2003) described how features of the material and social environment that people use to take decisions were also neglected in Clark's theory.

Since language is not a direct translation of a speaker or writer's knowledge, the interaction between knowledge and language that I find within CSCL and CSCW, is a complex one (Alamargot & Andriessen, 2002). While everyday human interaction has developed to be very efficient in the

[9] This section is partly based on a paper written by Cherubini et al., 2005.

recognition of mutual intentions, communicating about knowledge (defined as *semantic grounding* by Baker et al., 1999) does not rely on the same unproblematic and self-regulating character of *grounding-for-conversation*. Building a shared understanding requires a certain degree of meta-knoewledge, which is not driven by the same automatic processes of communication. My reason for stressing this, is that while I believe in the great potential of communication to produce learning, I want to caution that not all communication will automatically do so. When analysing or designing for collaborative learning, we need to take into account the idea that successful conversation is not necessarily the same as successful knowledge sharing.

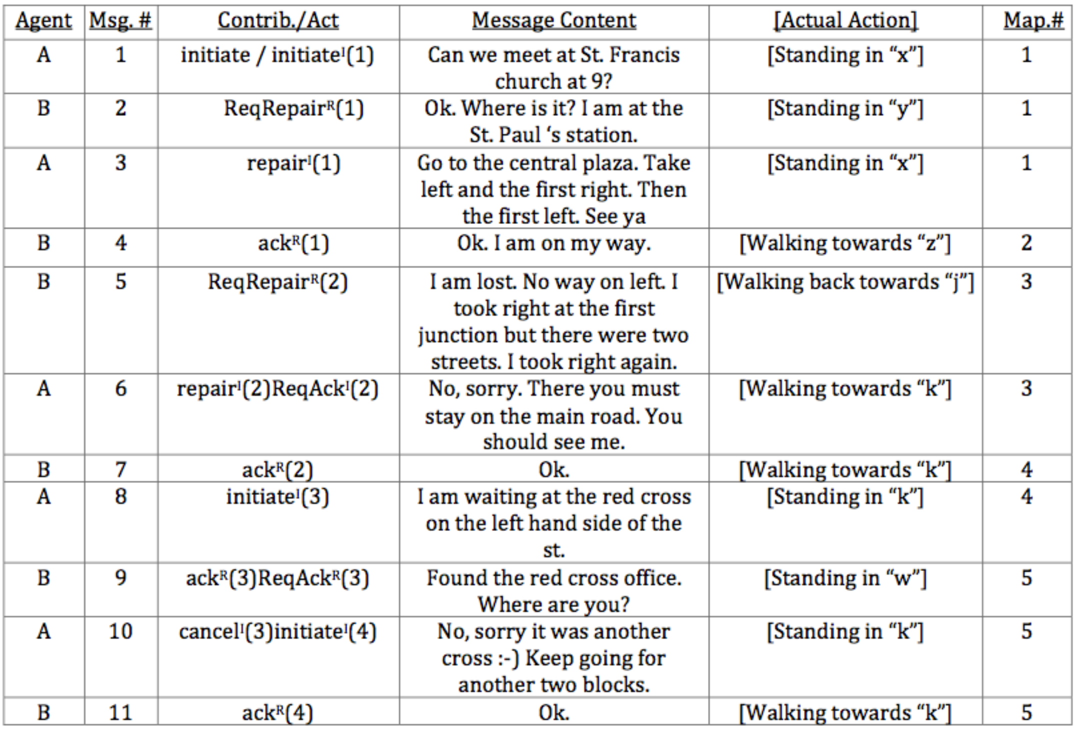
Below, the (subtle) differences of the characteristics, evidence, principles and mechanisms of grounding at the micro and macro level will be discussed. To give this a practical context, I will present an example drawn from the use of mobile messaging in a spatial collaboration task.

Example

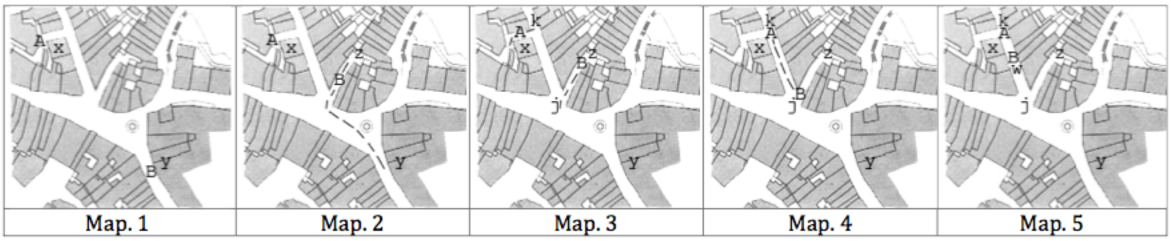
This example illustrates the limited information that acknowledgements give us about grounding at a knowledge level with an instance of human-to-human IT mediated communication, where two agents are coordinating for a meeting in an urban environment. The two peers exchange SMS messages containing directions and positions, with the aim to reach a physical co-presence. Below, I will report the exchange transcript and the references made to a city map (see table 2.4).

If I try to model the described situation using Clark and Shaefer's Contribution Model (1989), or Traum's Grounding Acts Model (1999), I reach the conclusion that `A` and `B` have grounded their conversation at each acknowledgment. More precisely: once both presentation and acceptance phases have been completed, the peers will have grounded a certain contribution (at utterance n. `4`, `7`, `9`, `11`). There is the tendency in CSCL and CSCW to correlate the rate of acknowledgment with the level of shared understanding on the assumption that the provision of evidence of reception is enough to infer the understanding of the signal and the corresponding incorporation in the contributor's beliefs. Additionally, when using these models, it is difficult to operationalise a lack of understanding, as in the example provided when `B` leads to point "`z`", because `B` provided clearly evidence of acceptance, as per message `4`, on table 2.4. My claim is that in order to take into account the complexity of this kind of interaction we need to look at the situation from a knowledge construction point of view. From there new descriptors of grounding are needed. Therefore, to continue with my example, we can say that the respondent `B` had an **illusion of grounding** between point "`y`" and point "`j`", until she realised that multiple solutions were possible and she did not have enough information to solve the ambiguity.

Table 2.4: Transcript of the example conversation. In the third column I coded the transcript using the formalisation proposed by Traum (1999). In the bottom part, map references used in the transcript



Agent	Msg. #	Contrib./Act	Message Content	[Actual Action]	Map.#
A	1	initiate / initiate^I(1)	Can we meet at St. Francis church at 9?	[Standing in "x"]	1
B	2	ReqRepair^R(1)	Ok. Where is it? I am at the St. Paul 's station.	[Standing in "y"]	1
A	3	repair^I(1)	Go to the central plaza. Take left and the first right. Then the first left. See ya	[Standing in "x"]	1
B	4	ack^R(1)	Ok. I am on my way.	[Walking towards "z"]	2
B	5	ReqRepair^R(2)	I am lost. No way on left. I took right at the first junction but there were two streets. I took right again.	[Walking back towards "j"]	3
A	6	repair^I(2)ReqAck^I(2)	No, sorry. There you must stay on the main road. You should see me.	[Walking towards "k"]	3
B	7	ack^R(2)	Ok.	[Walking towards "k"]	4
A	8	initiate^I(3)	I am waiting at the red cross on the left hand side of the st.	[Standing in "k"]	4
B	9	ack^R(3)ReqAck^R(3)	Found the red cross office. Where are you?	[Standing in "w"]	5
A	10	cancel^I(3)initiate^I(4)	No, sorry it was another cross :-) Keep going for another two blocks.	[Standing in "k"]	5
B	11	ack^R(4)	Ok.	[Walking towards "k"]	5



Map. 1 Map. 2 Map. 3 Map. 4 Map. 5

Model of grounding at a micro and macro levels

Using the presented example, I now elaborate on the difference between the micro and macro level, in four interrelated dimensions (see figure 2-14). Firstly, my examples show that the broad range of possible meanings on a knowledge level makes grounding more difficult, and is more likely to result in partial understanding then at a conversation level. Secondly, when it comes to measuring successful grounding, I propose to look at levels of commitment and co-referenced action, which might demonstrate (degrees of) shared knowledge better then acknowledgements. Thirdly, I will look at the underlying principles and see that because grounding is essentially efficiency-driven, the notion of 'effort' plays a central, but different, role at both levels. Finally, I will investigate where this effort is or should be directed and identify *perspective taking* (Järvelä & Häkkinen, 1999) as a primal grounding mechanisms on the knowledge level.

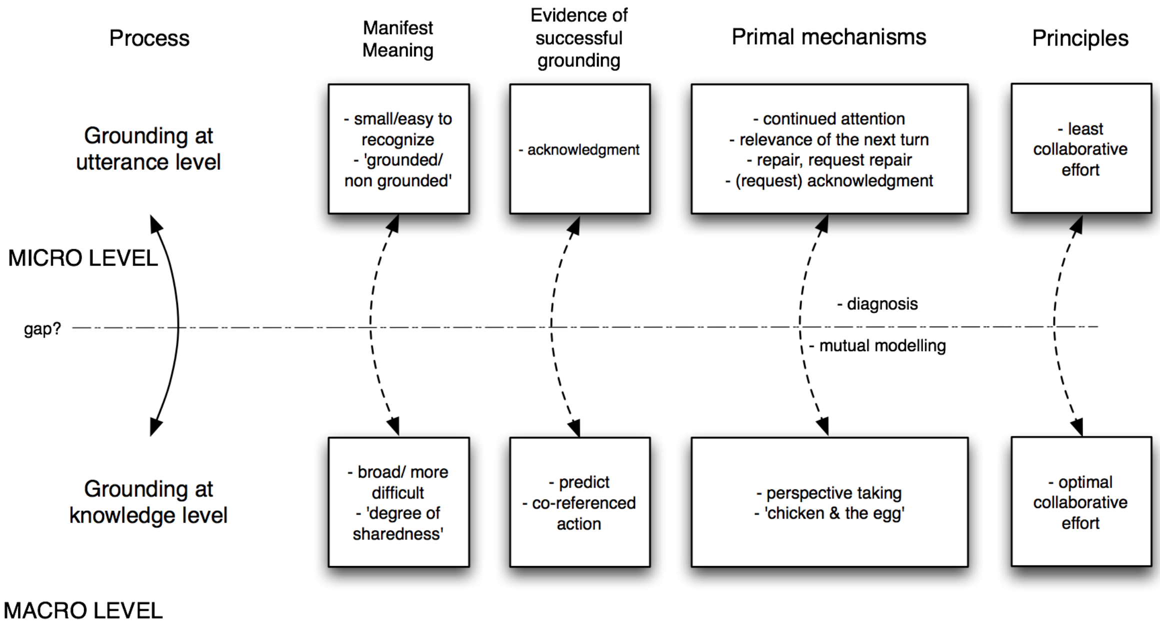


Figure 2-14: A four-component model of grounding at utterance and knowledge levels

Manifest meaning

Knowledge can never be accessed directly. As Laurillard (1993) states, I have to infer conceptual information (*descriptions of the world*) from the physical or communicative interactions I make in this world, thus making abstract learning, or communicating about knowledge, an essentially mediated phenomenon. Since this mediation is never perfect, and common ground can never be reached completely (Draper & Anderson, 1991, referring to Wittgenstein), I will use the notion of *mutual cognitive environment* instead (Sperber & Wilson, 1986). Sperber and Wilson define a cognitive environment as the set of facts that are 'manifest' at a certain moment to a person:

the facts that he or she is capable of representing and accepting as true or probably true. Or, in our words, what is manifest for a certain person is the range of possible meanings that are evoked or triggered by the evidence presented in a certain context. This collection of meanings that are associated to a certain action, concept or statement can even be so broad that it includes contradictory points of view (Bereiter, 2002). The difference with Clark's description of common ground lies in the fact that saying that two people share a cognitive environment does not imply they make the same assumptions; merely that they are capable of doing so.

While Clark's experiments started from the idea that a piece information `x` is either known or unknown to person `A` or `B`, the notion of *manifestness* shows that there are also many degrees in between, and many different ways of 'knowing piece of information `x`'. I can postulate that the bigger the overlap between the manifest meanings of different conversation partners, the more successful their grounding. When looking at the two levels I distinguished, I can state that the need for a notion of *groundedness*, which can account for subtle differences in interpretation, is even greater at a knowledge level than it is at an utterance level. Or, as Andriessen and Alargamot put it (2002, p. 8): "*semantic understanding is something gradual*". Also, the smaller and more focused a range of manifest meanings is, the better the chances for successful grounding. This depends on what one is grounding: an intention or speech act, a literal meaning, a statement, or a certain point of view. The more elaborate and complex the grounding object, the more difficult grounding. Because the range of possible interpretations will usually be broader at the knowledge level than at the utterance level, grounding will also be more difficult at that level, and "*a communicative intention can be fulfilled without the corresponding informative intention being fulfilled*" (Sperber & Wilson, 1986). The distinction I make in this section between the micro and the macro level should not be intended as a dichotomy, but rather as a range, for instance going from recognizing simple intentions, to recognizing literal meanings, more elaborated statements and finally complex points of view.

Evidence of successful grounding

In concordance with Sperber and Wilson's account of the evidence that messages provide to guide their interpretation, the same can be said about analysing grounding. The more evidence we have, the more we know about the levels of shared understanding (though it may never be conclusive). As I have stated in the introduction, I do not think acknowledgements are always a valid measure of shared understanding. Apart from different goals at the two levels (see the grounding principles at page 53), Ross et al. (1977) have shown that an (partial) *illusion of shared knowledge* is not only possible, but also even likely to occur (called the *false consensus effect*). Bereiter's term *knowledgeability* (2002), or 'being able to take intelligent action', indicates that (verbal or physical) actions intrinsically contain knowledge. In my example (section 2.4.3) the 'information bearing actions' one can identify are the coordination of tuning attempts with the agreed plan. If the pair agrees to a certain strategy and then implements it coherently, we can infer that the pair successfully grounded to a high degree. Or, more generally, if someone "commits to a previous statement, and subsequently does something directly related to it in the forthcoming action or statement" (we use this notion of commitment in accord with DiEugenio, Jordan, Thomason, & Moore, 2000) then the statement was grounded at knowledge level. Since this relatedness between communicative actions requires a large overlap in the cognitive context and shared referents, I will label them as *co-referenced actions*.

In asynchronous discussion groups (e.g, fora) we can look at the alignment of questions and answers. An answer that follows a question might seem like a legitimate and useful speech act (utterance level). We can deduce whether it is also a successful knowledge-building act only if the relevance of the content is established. On a knowledge level, for an action to be *co-referenced*, it needs to refer to a shared piece of knowledge and needs to be relevant from someone else's view. According to Sperber and Wilson (1986, p. 608):

> ... something is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him: say, by answering a question he had in mind, improving his knowledge on a certain topic, settling a doubt, confirming a suspicion, or correcting a mistaken impression.

The example above shows that, while at an utterance level, both recognising a certain speech act, (such as identify a question by its question mark) and providing a relevant response (giving an answer) is pretty straightforward, on a knowledge level, the requirements for action to be relevant or co-referenced are much higher.

Grounding mechanism

At an utterance level, human communication is very efficient as minimal effort is invested in message design and interpretation (by jumping to –subjective– conclusions and repairing a possible misunderstanding after it arises). At a knowledge level however, I have argued that because of the mediated nature of grounding and the more complex collections of associated (manifest) meanings, this efficiency presents more problems. Miscommunication can be both harder to detect (thus cannot be relied upon to reveal itself) and to repair. Therefore, the grounding mechanisms at the knowledge level might present the most important shift from the utterance level. To understand what nuanced meaning other people attribute to certain statements, one must "put oneself in the other's shoes" and try to identify which meaning will be relevant for that person (Sperber & Wilson, 1986). In order to infer someone else's cognitive environment or *frame of reference*, both for reading and writing messages (audience design), we rely on strategies like perspective taking (Järvelä & Häkkinen, 1999) and mutual modeling (for a definition see Nova et al., 2003).

While at an utterance level, repair mechanisms are know to be self-regulating (the less shared understanding, the more grounding will take place, see for example Dillenbourg & Traum, 2006), this is less evident for mechanisms like perspective taking, happening at work, at the knowledge level. It seems that at this level, there is a 'chicken and the egg' relation between grounding and common ground ("*It is hard to find some if you don't have some already and you don't have any unless you find it*", Krauss & Fussell, 1990, p. 4) is even more prevalent than it is at the utterance level. This shows that at a macro level, knowledge of other's perspectives plays a role as a prerequisite as well as an outcome and the same goes for one's knowledge of the subject matter. Because identifying another's frame of reference is easier if one has knowledge of the different possible frames of reference that exist, perspective taking is also tied to existing knowledge. This underlines the reciprocal relationship between individual and collective processes in collaborative work or learning (Stahl, 2000): it is not only that individual learning results from collaborative processes, but also that individual knowledge also influences the success of collaboration.

Grounding principles

First of all, grounding is functional and driven by mechanisms of efficiency, as Clark and Wilkes-Gibbs (1986) demonstrate with their *principle of least collaborative effort* and Wilson and Sperber (1986) in their *relevance-theoretic comprehension procedure*. The fact that in grounding no more effort will be invested that what is 'sufficient', can explain the lack of co-referenced actions in my example. For collaborators the costs (relative to the goals) may simply be too high, especially because high-level collaboration goals are usually translated into practical tasks, with which collaborators deal in a pragmatic way. Taking the perspective of someone else may take more effort than staying within one's own perspective, and what is 'sufficient to continue the conversation' might not be sufficient for collaboration (Baker et al., 1999). That is why, for collaborative work, instead of trying to 'minimize the collaborative effort', we strive for an: *optimal collaborative effort* (Dillenbourg, Traum, & Schneider, 1996).

The effect of effort into perspective taking and co-referenced actions is twofold: not only does relevant feedback enhance collaborative knowledge building, but the effort of creating shared meaning itself is also strongly associated with learning (Schwartz & Lin, 2000), especially if the effort is directed at the knowledge level (or 'semantic grounding', see Baker et al., 1999). Spending effort in trying to understand another perspective is learning: it is leaving behind one's preconceptions and exploring new information and insights. The is also true for reading, since perspective taking for comprehending messages is closely related to the comprehending process when studying scientific texts.

To conclude

Context is inextricably present when we grasp meanings and when we infer knowledge. The pragmatic tradition of relevance highlights the action-oriented nature of intelligence, where the term 'action' is to be understood in a broad sense that includes reasoning behaviours, or communicative acts (Ekbia & Maguitman, 2001). When looking at the relevance of communicative actions in collaborative learning, I have described how providing evidence and acting in a co-referenced way is crucial for developing a shared understanding. The more evidence is presented, the easier it becomes to take another's perspective, act in a co-referenced way and enhance the degree of shared understanding. I suggest the implications for design and research on grounding in collaborative learning might involve an effort to facilitate grounding at a knowledge level. For instance, communication tools could be developed that provide more (focused and detailed) contextual information which serves to limit the range of manifest meanings of the concepts that are being used, and thus to increase the chances of shared understanding. Tools implementing Explicit Referencing follow this principle. Also, since the use of acknowledgements as markers of shared understanding is problematic, I propose to create markers that can give an account of the relevance of communicative actions in regards to the reasoning process. As an example of this, this chapter argued that making operational the 'co-referencedness' of actions on a knowledge level, as measure of shared understanding, would be a valuable effort.

2.4.4 Synthesis: research framework

The beginning of this section introduced the grounding theory developed by Clark and colleagues. During the years, some criticisms have been raised against this theory and its ability to model human communication. Some of these criticisms cannot be dismissed. Other arguments carried out by scholars such as Sperber and Wilson on the notion of common ground can be countered by arguing that they refer to early versions of Clark's work. In his late papers, he reconsiders the importance of cognitive, situated and social aspects of interactions, in the establishment and maintenance of common ground. More recent critics on the *degree of consciousness* of linguistic mechanisms such that of common ground are recurring problems in pragmatics. The resolution of these debates is outside the scope of this thesis.

What is of interest for this work is the connection of space and communication, which I will interpret as a coordination device. Collaborators trying to reach a public goal will face inevitably a coordination problem. Solving the problem requires the participants to figure out how to contribute to the joint action by inferring the intentions of the other participants. This process is made possible by the exchange of coordination devices, which contributes to the common ground. Also, I describe this mechanism as *situated* in that it does not happen constantly but it is related to problematic situations. In this work, I am less interested in the inference mechanisms per se. Rather, I am interested in how a coordination device influences communication processes.

Finally, I will use Clark's theory to understand mechanisms of communication happening at *knowledge level* rather than linguistic level. The last section detailed how these two levels present differences of processes and related observables. As argued, the micro or linguistic level is not directly related to the establishment of common knowledge necessary to solve a problem collaboratively. This is why I did not enter into finer level details of Clark's theory.

2.5 Conclusion

It is worth to conclude this chapter by reviewing the most important theoretical points that explain the connections of space, language, and interaction.

– *Language structures space and space structures language.* The way spatial relations are expressed is typical of a language. Language is incredibly efficient in capturing the vast number of possible configurations of reality that might be needed to communicate. It reaches this efficiency with a limited number of words that can assume a variety of meanings and despite a multitude of misunderstandings that can arise from their use. In fact, people are good in assigning the right meaning to these expressions through a series of cognitive and linguistic strategies. One of this strategy is the binding of language and the contextual space from which it refers (e.g., the use of a deictic gesture). Given a certain situation only a limited number of interpretations are admissible. The other way around: only a limited number of linguistic expressions can encode correctly the given –contextualized– situation. Spatial situation are schematized in mind, encoding their relevant features. This cognitive map has itself spatial features that influence how people think and describe its content.

– *Language is embodied.* The way meaning is inferred from linguistic expressions can not be understood only in terms of informative content expressed by the linguistic items. Extra-linguistic constraints need to be taken into account to understand how people assign meaning to expressions, either when they encode meaning into utterances, than when they decode this meaning from messages they receive. Spatial language can only be understood by taking into account extra-geometric knowledge of the object involved in the spatial scene to be described. For instance, it is unusual to say that *the head is in the hat* because of our knowledge of gravity and mass of these two objects and because of our attribution of will to the head instead of the cloth covering it. Gestures are prototypical communicative devices that embody language binding space and linguistic expressions. People's speech falters when they are asked to describe space without using their hands so as to prevent gesturing (Rauscher et al., 1996).

- *Maps condense the interplay of language and space.* The use of maps has been proven to have positive effects on cognition and learning of spatial situations. They communicate many concepts of an actual space through the layout and forms of their graphical elements. While these artifacts have been studied extensively for their ability to support geographical coordination in ubiquitous situations the possibility of sharing map annotations received much less attention by the research community. On the other hand, many informal applications of map annotations are used fruitfully by different communities. The understanding of how language is used in conjunction with maps may yield many interesting applications and further our understanding of the interconnections of language and space.

- *Maps annotations are worth investigating.* While maps annotations are widespread and used in a multitude of formal and informal situations, little research has explained whether and how digital maps annotations could be used for sustaining collaborative work at distance. Since I saw an increased development of systems combining maps and social interactions, I needed a deeper understanding of how enriching messages with location compares to a non-contextualized communication.

- *Mobility and text communication.* Ethnographical observations of use of mobile phones revealed a widely diffused practice of sharing SMS to maintain location awareness among group members. Mobile technologies, in particular, make evident the need of remote conversants to reconstruct the context of interaction that has been fractured by distance. Knowing the location of the conversational partner is extremely useful in inferring the person's activities, availability for further interaction and so forth. While it is clear the importance of this form of communication in informal situations, little research has focused on the use of textual information and mobility for supporting formal knowledge representation and sharing. Whether the connection of communication and positioning information could support remote collaboration and the design of useful application has yet to be demonstrated.

The last section of this chapter presented the psycholinguistic framework that will be used to analyse the results of the experiments presented in this thesis. This framework is based on the grounding theory developed by Clark and colleagues. However, I argued that for the study of collaborative interactions, existing theories of grounding such as that of Clark and Shaefer (1989) cannot be applied without adjustments. When comparing collaborative work and conversations, four dimensions can be identified where grounding at a knowledge level differs from the grounding at an utterance level. Firstly, the indirect access and the existence of a range of manifest meanings, poses the need for a notion of *groundedness*. Secondly, I proposed providing evidence in *co-referenced actions* to be an important process as well as an additional marker to assess grounding. Thirdly, instead of simply repairing misunderstandings after they arise, *perspective taking* becomes a more prominent mechanism. Fourthly, effort put into grounding is shifted from needing to be minimised, into needing to be *optimised*.

Chapter 3

Deixis in Dual Spaces

This chapter reviews previous research in the area of remote support for deictic gestures. First, a framework for evaluating CSCW applications will be introduced, then the chapter will describe tools that enable remote gestures and the related discussion on studies that have been performed with them.

3.1 Deixis as a basic pragmatic need of collaboration

To discuss previous work that I considered relevant for this research, I will adapt a simple framework that has been used to analyze groupware (Dix, 1995). Any collaborative working situation presupposes different participants involved. These are denoted by circles labeled 'P' in figure 3-1. Collaboration also presupposes communication between the participants, which is represented by the arrow between the circles. Communication can be conceptualized as a series of units or messages. These are represented in figure 3-1 by the single circle 'M'. The aim of the collaboration will probably require a manipulation of elements in the shared visual space. These objects, which may be located in different places and controlled by different actors, are defined in the framework as 'artifacts of work' and are represented in figure 3-1 by the single circle 'A'.

Research in supporting collaborative work at a distance can be organized into groups according to two distinct strategies: studies striving to replicate face-to-face settings, and those leveraging on an increased awareness and coordination of the peers through a shared visual space. The first approach is represented in diagram 3-1 by the path P–M–P and will be detailed in section 3.2, while the second approach is represented by the path P–A–P and will be detailed in section 3.3. As I will explain later on, many years of research have revealed how we are still far from developing technology that allows people to collaborate at distance with the same efficiency that we have when face-to-face. Ethnographical observations from real work settings show how

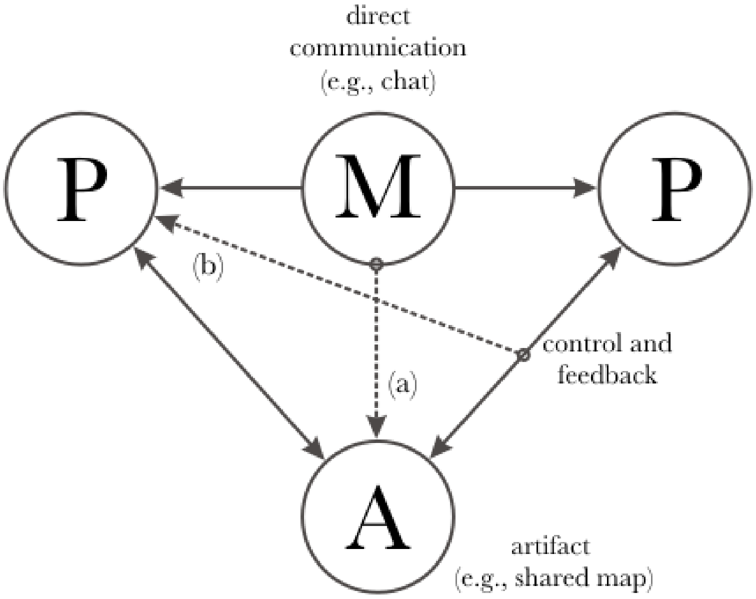


Figure 3-1: Framework of Computer Supported Collaborative Work. (a) deixis or explicit referencing (see section 3.5); (b) feedthrough (see section 3.3.2)

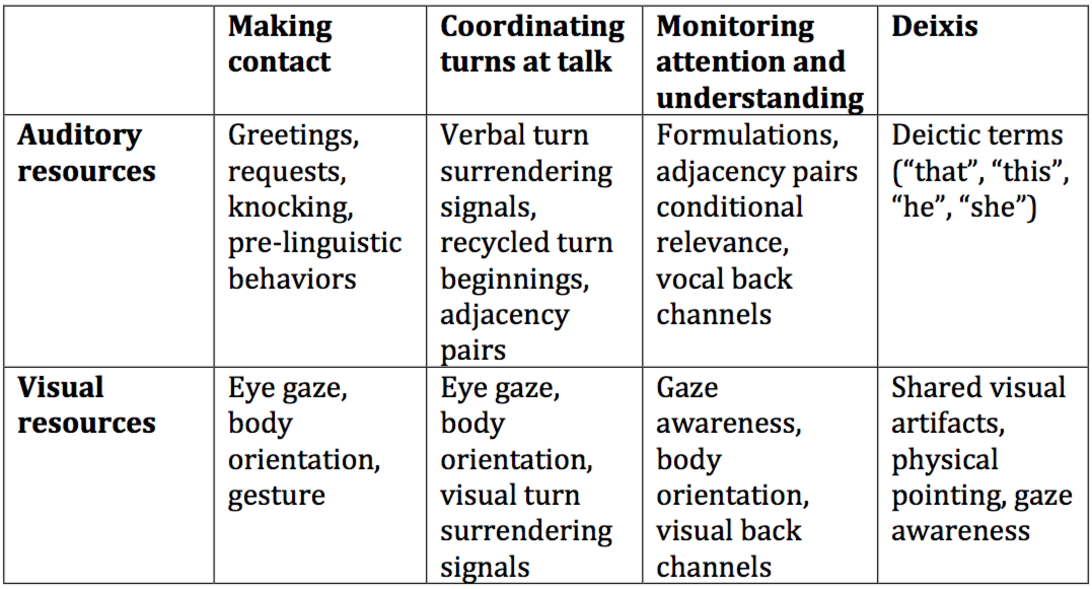
many solutions developed to support collaborative work at a distance following the above two approaches above flawed as they "*fracture the relation between action and the relevant environment*". For example, using many video cameras to capture and share different points of view between two remote locations might seem an improvement over the use of a single camera. However, users might feel lost in the attempt to understand which view is the partner currently using or how to adapt common communication strategies to this multitude of perspectives. As Luff et al. explain (2003, p. 73, my Italic):

> In developing systems to support remote collaboration, we attempt to interweave or create new environments in which participants can produce actions within a framework or setting, which, in part, is accessible and intelligible to each other. However, once we begin to create new environments to enable people to interact and collaborate with each other, we fracture the relation between action and the relevant environment and thereby engender difficulties that may render even the most seemingly simple form of activity problematic. Ironically, *the more we attempt to enhance the environment, the more we may exacerbate difficulties for the participants* themselves in the production and coordination of action.

In this regard, many critics have been raised against the assumption that the best environment for computer-supported collaborative learning or work is the one that most closely reproduces the feature of face-to-face collaboration. This has been defined as *face-to-face imitation bias* (Dillenbourg, 2003). Therefore, we need a different approach to deal with the problem. We need to find more subtle technological solutions to translate communication mechanism which are effective in presence but not available when collaborators are not co-located. These solutions should allow to recreate the same functions using different but equivalent strategies. The aim of this thesis is to focus on one of these mechanisms, namely deixis (represented as the path M–A in figure 3-1, and detailed in section 3.5), to understand how this process can be enabled in virtual environments and how its different implementations may impact collaboration.

Deixis is often described as a basic pragmatic need of collaboration. Effective collaboration requires that participants can efficiently communicate their intents, agree on a methodology to achieve their goals, share information and monitor the development of the interaction. Daly-Jones et al. (1998) defined four pragmatic needs that must be fulfilled through the transmission of either auditory or visual information in human interaction: (1) the need to make contact; (2) the need to allocate turns for talking; (3) the need to monitor understanding and audience attention; and finally (4) *the need to support deixis*. These points are summarized in table 3.1, where possible roles for auditory and visual information are considered.

Table 3.1: Resources fulfilling four roles for auditory and visual information in human interaction, from (Daly-Jones et al., 1998, p. 23)



	Making contact	**Coordinating turns at talk**	**Monitoring attention and understanding**	**Deixis**
Auditory resources	Greetings, requests, knocking, pre-linguistic behaviors	Verbal turn surrendering signals, recycled turn beginnings, adjacency pairs	Formulations, adjacency pairs conditional relevance, vocal back channels	Deictic terms ("that", "this", "he", "she")
Visual resources	Eye gaze, body orientation, gesture	Eye gaze, body orientation, visual turn surrendering signals	Gaze awareness, body orientation, visual back channels	Shared visual artifacts, physical pointing, gaze awareness

The last two points are particularly interesting for the design of systems for remote collaboration and the related literature presented in this section. When we communicate the production of our elocution is inextricably linked to the responses of our audience. It is crucial for the speaker to monitor his or her audience for evidence of continued attention and understanding (Clark & Brennan, 1991). Argyle (1969) showed how gaze patterns, facial expressions, gestures and body posture are all mechanisms used for collecting evidence regarding the level of attention, attitude and response of an audience. However, attention does not equate with understanding or agreement. Many strategies are at play in human communication. For instance, conversants can offer explicit verifications of the interpreted content of previous utterances. Or, an utterance that serves as a request for information may be deemed to have been interpreted successfully if a relevant answer is given. This information is given or gained though vocal backchannels or visual ones, like gestures. Pointing is one of the actions that a speaker might perform in order to make her utterances as intelligible as possible. McCarthy and Monk (1994) argue that the ability to point to a shared artifact accounts for the extreme efficiency of many utterances in everyday communication. Deixis can be produced through the auditory channel using specific terms like "this", "he", or "here", taking advantage of the recentness of previous uttered information. In this case I will talk of anaphora, or linguistic deixis. Otherwise, deixis can be produced through multimodal signaling, as an interaction of linguistic messages and gestural pointing (McNeill & Levy, 1982). In this last case, deixis ties a message to a location or an object in the shared visual space. This shows how *indicating* is inseparable from the associated communication (Heath et al., 2001), but more importantly, it shows how deixis is most effective when the "talk" can be clearly separated from what is talked about. Studies like that of Whittaker et al. (1991) or Dillenbourg and Traum (2006), show that when remote collaborators can choose among a variety of communication channels to structure their interaction, the most persistent medium is preferred to construct a shared artifact to talk about, while the talk itself is transferred to the less persistent channel. This is consistent with the idea of Hutchins and Klausen (1991), who suggest that the artifacts used in cooperative tasks support "distributed cognition", in that they allow people to externalize their thought processes so that they can be shared by other members of the group. Finally, deixis is intertwined with gaze. The emitter of a message containing a deictic term uses sight to monitor the attention and understanding of the produced deixis itself. The receiver also uses sight to associate the visual context to the message.

This discussion summarizes the importance of deixis in collaborative work. It also introduces some of the issues that will be discussed in more detail throughout the rest of this chapter and that hindered the success of many systems developed in the last years to support collaborative work at a distance. I will argue that many issues that affected these systems are connected with how: (a) the importance of deixis was neglected; (b) its implementation was under- or over-implemented; (c) deixis was not properly connected with communicational intentions; and finally (d) it was not properly backchannelled with gaze.

The rest of this chapter will develop these points following a loosely chronological list of the systems supporting remote collaboration through the last four decades, first by enhancing the connection of the participants with better video links (section 3.2), then by ameliorating their ability to have object-focused discussions (section 3.3), by more specifically allowing for remote gestures (section 3.4), more specifically remote deixis (section 3.5), and finally through specific solutions for supporting gaze awareness (section 3.7).

3.2 Video-Mediated communication

In this section, I will look briefly at research aimed at sustaining collaborative work at a distance through Video-Mediated Communication (VMC), an integral aspect of most remote gesture tools (Kirk, 2006). Research on VMC started more than three decades ago (the earliest studies on the subject were conducted by Chapanis and colleagues, 1972) and aimed at providing collaborators visual access to a remote site. It is not my purpose here to make an extensive revue of the evolution of VMC technologies, but only to show some phases of the research in this field that are relevant for my research questions. VMC tools evolved over the years by following a face-to-face imitation paradigm and increasing system features. Eventually, this led to inconclusive findings about the efficacy of these systems to support group work. A more detailed overview of the research in the field is provided by Finn (1997) and Kirk (2006).

3.2.1 Technologies for VMC

Earlier work on VMC focused on videoconferencing systems which sought to support formal meetings. Examples of such technologies included systems such as ISDN and LiveNet, discussed by O'Conaill and Whittaker (1997), and the video conferencing systems described by Tang and Isaacs (1993). Such systems were adopted by large corporations and employed sometimes ad-hoc wiring between remote sites. The main characteristic of these systems was that they were designed for group meetings using a large screen monitor and a single camera held above the monitor, as for instance the XTV system described by Sellen and Harper (1997).

Later work considered more informal settings aimed at supporting desktop videoconferencing (Tang and Isaacs, 1993, Sellen, 1995). The main objective was to provide video-based access to multiple participants working at different locations. Most of these system employed a strategy of Picture-In-Picture (PIP) presentation, where only the torso of each participant was visible (see figure 3-2). These systems were located close or embedded into PC system, and soon prototypes began to incorporate document sharing capabilities. Subsequently, these systems began supporting object-focused conversation instead of simply allowing for remote meetings.

Subsequent development in VMC technology continued to incorporate informal aspects of day to day communication into remote collaboration systems. Media Spaces integrated video connectivity into the architecture of working spaces, with easy to reconfigure video links between remote locations (see for instance figure 3-3, the media space developed at Xerox PARC, Bly et al.,

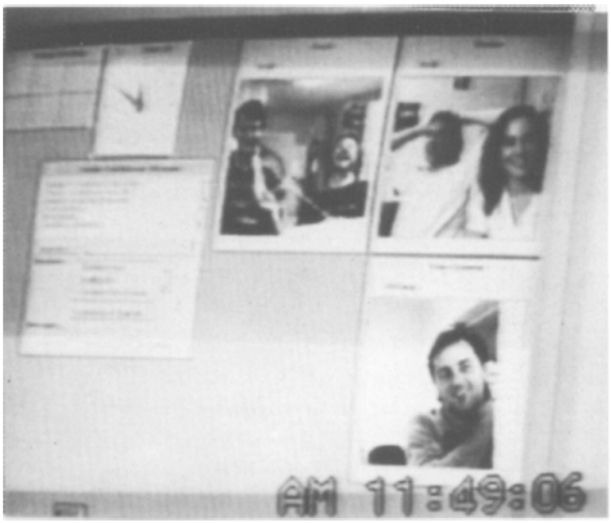


Figure 3-2: the DVC prototype of Tang and Isaacs, (1993)

1993). A media space was defined as a networked computing environment that is never turned off. The design of these systems underwent many iterations as users expressed different needs as accessible file repositories between remote locations or simple video feeds used for informal glances.

However, media spaces have not reached the mass market, because they required an expensive infrastructure to work, and were limited in the scope of access to the remote site that they could provide. Also because these 'always on' systems exposed the privacy of the personal workspace of the users and finally because there was no striking evidence of benefits for remote collaborators.

A reconsideration of the design assumptions for VMC technology arose out of the first systematic ethnographical studies of these systems in the early 90s. Heath and Luff (1991) observed researchers using video-mediated communication systems at EuroPARC. The authors described some problematic aspects of VMC: gestures were rendered ineffective by technology as their motion and perception was disrupted by the fixed and narrow filter of the video. Also, actions and body movements were ineffectual in securing gaze and therefore users abandoned their adoption in favour of other strategies of communication.

Researchers designed systems allowing the users to switch between multiple cameras in order to provide more flexible access to remote working environments. One of the best examples of such a system was called Multiple Target Video (MTV, see figure 3-4). Gaver et al. (1993) conducted observations of collaborators using this system to understand how, and why, people switched among multiple cameras. Their system allowed to switch between multiple cameras. One was focusing on the shared workspace. Another was looking at settings of the remote collaborator's office from a bird's-eye perspective a third was installed on each remote desks offering capturing

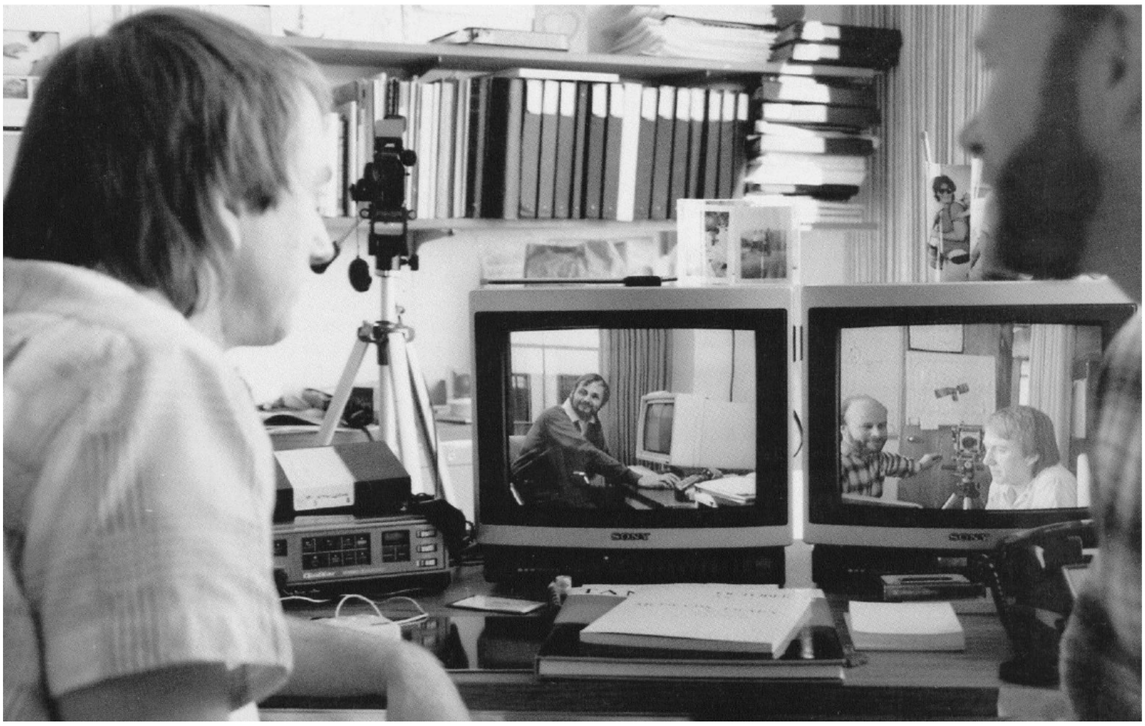


Figure 3-3: A media space. The remote location is visualized on the screen. Also the screens used for displaying the video feed of the meeting are distinct from the computer screen (from Bly et al., 1993)

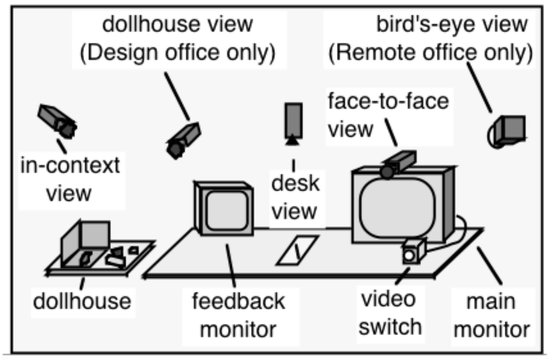


Figure 3-4: The MTV set-up consisted of several cameras. The user could select which remote camera to show on the displays (from Gaver et al., 1993)

the torso of the collaborator. Additionally, it was possible to add extra cameras for specific requirements. However, the resultant views were of minimal importance to the task. The view of the torso of the remote collaborator was not used intensively. Additionally, the authors found two big limitations of the system that impeded an efficient use towards a furniture arrangement task. First, a collaborator could not point to objects in the shared workspace. Second, a collaborator could not attract the attention (e.g., the gaze) of the other collaborator. In sum, MTV systems neglected the importance of the context in which the collaboration was happening in favour of the default view of the torso of the remote collaborators. The major problems of this technology was that looks passed unnoticed, gestures were impotent and resultant presentations were distorted and incongruent (Heath et al., 1997).

The same results were reported by Hindmarsh et al. (1998). They observed that one significant limitation of such interaction spaces was that individuals could not easily determine to what or whom a participant was referring. The problem arose from the difficulty in connecting an image of the other with the image of the object she was referring to. The authors noticed that in these environments, object-focused discussions are problematic because of the 'fragmentation' of different elements of the workspace. In co-present interaction, participants can usually see fellow participants and relevant objects in relation to their surroundings. In interactions at a distance participants have to re-assemble the relations between the body and the object. Participants observed by Hindmarsh and colleagues tended to overcome these limitations making implicit references more explicit. For instance, instead of saying: "what do you know about this" they would say: "See this sofa here?". The authors noted that the major problems of this technology were a limited horizontal field of view, a lack of information about others' actions, slow movements, and a lack of support for executing multiple actions at the same time.

3.2.2 Experimental evidences of the impact of VMC on remote collaboration

Many studies of VMC aimed to establish its efficacy in measurable ways. Earlier studies focusing on the task performance, did not establish that there were performance enhancements from the provision of video links between remote collaborators. Studies conducted by Chapanis and colleagues (1972) concluded that the audio channel was crucial for collaboration. The same finding was confirmed by later comparisons of different media configurations supporting interaction (Minneman & Bly, 1991; Olson et al., 1995; Whittaker, 1995).

However, when research focused on the communication process, the experimental comparisons of VMC versus face-to-face interaction demonstrated that the use of VMC led to more formalized turn-taking, and fewer interruptions (O'Conaill & Whittaker, 1997). Even if VMC failed to replicate the quality and efficiency of face-to-face meetings, these studies proved that high quality VMC improved the process of communication, making it more similar to co-located interaction than audio-only communication technology. This finding was confirmed by the study of Sellen on different forms of VMC (Sellen, 1995). She compared different forms of VMC with face-to-face and audio only conversations. Sellen noticed higher levels of interruptions in face-to-face and she argued that this was an indicator of interactivity, which she defined as *fluent interactions*. Both Sellen and Anderson et al. (1997) noticed that VMC leaded to more interruptions than audio-only interactions. However, they noticed how improving VMC connections with full eye contact did not make interaction the same as face-to-face communication.

Some researchers compared different communication conditions in an experimental settings where they included a shared editing tool for a collaborative task, and measured conversational fluency and interpersonal awareness. Olson et al. (1995), for instance, compared study groups of three people designing an Automated Post Office. They used three conditions: high-quality video links between the remote sites, audio-only links, and a third situation in which participants could use a shared editor to support their emerging design. Participants produced higher-quality design when they could use a shared editor to support their collaboration. Even if participant rated an audio-only condition as having the lower communication-quality, the quality of work suffered little compared to a video communication condition. Similar results were obtained by Daly-Jones et al. (1998).

In conclusion, these results showed that task outcomes were not affected by the use of VMC technologies. An important point raised by researchers was that video led to more fluent and informal conversations when compared to audio-only technologies. However, when VMC solutions were tested in problem-solving situations, it was clear that the ability of looking at the face of the remote collaborator did not have significant influence on the performance of solving the task. The possibility of looking a the artifact used to solve the task was reported to have positive effects on performance. VMC technology did not seem to greatly enhance remote collaboration. Nevertheless, the availability of a shared visual space began to attract the interest of researchers as being implicated in the mechanisms of attention and conflict resolution. I will develop further this point in the next section.

3.3 Shared visual space

While some researchers tried to improve the efficacy of remote collaboration by the means of a better video link between the participants, another part of the community looked at how better performances could have been reached by providing a shared visual space (in short SVS) for collaborative tasks. This is an area of an interface whose content and presentation is maintained equal across different remote sites. Early work in this area was conducted by Krauss and Fussel (1990, 1991) with an experimental design aiming at exploring the process of achieving grounded conversations through different communication technologies. Kirk explains their initial key findings (2006, p. 27):

> Through their experimental analyses Krauss and Fussell began to understand how task-focused language evolved during its interactive use during collaborative tasks. The evolution of referring expressions and the developing awareness of common referents was demonstrably shown to be significantly effected by the resources used to

> establish communications. If a shared visual environment was enabled it was often observed to be of significant support to the smooth establishment of such critical communicative processes.

Fussell et al. (2000) demonstrated that video-communication technology was inadequate to establish shared visual spaces. Further, Karsenty (1999) expanded this argument saying that to support any given task it was crucial to determine which features of the visual environment were critical to support.

Gergle contributed to the comprehension of the impact of a shared visual space on collaborative work at a distance (Gergle, 2006). He designed a series of experiments around a puzzle task paradigm in which a Helper guided the actions of a Worker in a collaborative effort to assembly a puzzle piece diagram (see figure 3-5). Using this design Gergle, Kraut and Fussell (2002; 2004b) demonstrated that the presence of a shared visual space improved performance on the task. Additionally, Gergle et al. (2004) demonstrated that when communication is mediated by text-based chat, the persistence of the conversation improves performance as collaborators can make actions in parallel, thus economising time. However, this improvement of performance is smaller than the improvement that can be recorded when a shared visual space is available to the collaborators.

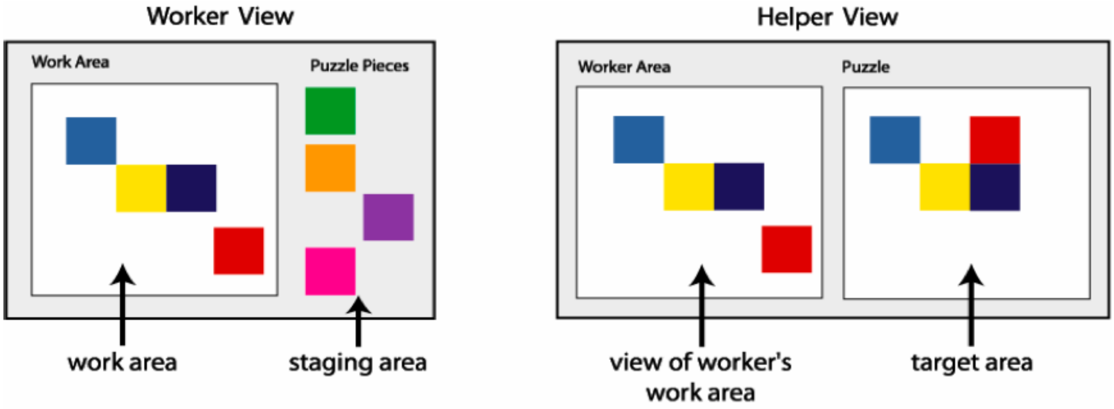


Figure 3-5: The collaborative puzzle task. The Worker's view (left) and the Helper's view (right) from Gergle (2006)

Using a sequential analysis, Gergle and colleagues further demonstrated how visible actions within the shared workspace can be used to replace utterances of the dialogue that would be necessary in the absence of visual feedback (Gergle, Kraut, & Fussell, 2004a). When participants have access to a shared visual space, they can visually monitor the evolution of the task and the comprehension of instructions exchanged. Therefore the resulting conversation will be free of the explicit checking and confirming usually carried out through the linguistic channel. In further work, the same authors experimented with the effect of visual delay on the synchronization of the shared space on the puzzle task. They found that the impact of delayed synchronization of the SVS was a function of the complexity of the visual environment (Gergle et al., 2006) (e.g., the number and shape of tiles in the puzzle).

The idea of using a shared visual space as an effective interaction mechanism started in the early 1980s with the work of Schneiderman, described in the next section. More recent work also reconsidered the possibility that the simple knowledge that collaborators were manipulating elements of the workspace might be useful to improve remote collaboration. This mechanism was called *feedthrough*, and it is marked on diagram 3-1 as an arrow connecting one of the participants to the action of the other over the shared space. I will detail relevant work on this mechanism in section 3.3.2.

3.3.1 Direct Manipulation

Direct Manipulation was originally defined by Schneiderman (1982, 1983) as a class of systems with a graphical interface that allowed them to be operated 'directly' using manual actions rather than typed commands (Frohlich, 1997) (see figure 3-6). This initial definition was later revised and complemented by the work of Hutchins et al. (1986), in which they defined the directness of an interface as the sum of *the engagement* and of *the distance* that the interface offers. Engagement refers to the locus of control of action within the system, while distance refers to the mental effort required to translate goals into actions at the interface and then to evaluate their effects.

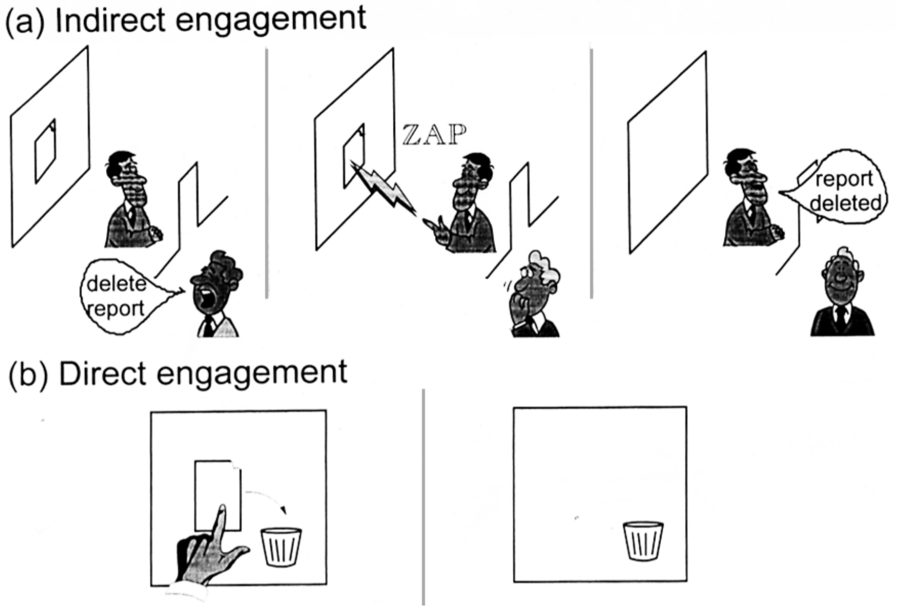


Figure 3-6: Two types of user's engagement (from Frohlich, 1997)

While there are debates on whether direct manipulation always leads to better performances at the individual level, we still know little of how this concept applies at the collaborative level. We can imagine that a shared visual space might be analogous of direct manipulation during collaborative tasks; however, collaboration requires the construction and maintenance of a shared representation of the problem (Rochelle & Teasley, 1995), which implies that the locus of control is divided between the collaborators, making manipulation essentially a mediation process. As noted by Frohlich (1997), the same kind of mediated manipulation of certain interfaces for which the user has to deal with an intermediary to get instructions executed (e.g., command line interfaces) happens in a different way in computer mediated communication tools, for which the interface agents are other people who may act on a shared workspace or document while talking to you (Whittaker et al., 1993).

3.3.2 Feedthrough and group awareness

Feedthrough is the term defined by Dix to describe the feedback that is offered when one of the participants in a remote collaboration acts on the artifacts of the workspace and this action is then visible to the other participant(s) (Dix, 1995). Feedthrough can be seen as a form of communication through the artifact (e.g., Mary asks Paul to open the window. Paul does not answer, however he opens the window. Mary does not need any verbal acknowledgment of the reception of her message). One of the most common forms of feedthrough embedded in chat applications, is the information of when a conversation partner is typing. Feedthrough is an active form of communication, to be distinguished from group awareness, which is a passive cognitive process.

Group awareness was studied by Gutwin and Greenberg in several works developed in the general framework of *workspace awareness* (1999, 2004). When people work together in a co-located setting, they can keep track of what the others are doing using a wide variety of perceptual cues. These are sources of information that stay in the cognitive background of people and become salient only when needed (e.g., Marc is typing a letter in his open space. Mary, the secretary is talking to John when she pronounces Marc's name. Suddenly Marc stops working on his letter and looks toward Mary to check whether she need something). Gutwin and Greenberg defined this awareness of others in the workplace as workspace awareness and demonstrated how the availability of this information is beneficial to sustain collaborative work.

3.3.3 Synthesis: design issues of SVS

The studies described in this section each demonstrate the positive effects of the availability of a shared visual space for remote collaborators (particularly Bly, 1988; Dourish & Bellotti, 1992;

J. C. Tang, 1991; Tatar et al., 1991; Whittaker et al., 1991, 1993). The empirical findings of these studies are consistent with linguistic theories that explain how collaborators should share a linguistic channel through which to exchange messages as well as the context necessary to interpret and disambiguate the literal content of any message (Clark & Marshall, 1981; Krauss & Fussell, 1990; Sperber & Wilson, 1986). A shared context is useful in reducing ambiguities and avoiding misunderstandings. On this base, it is natural to assume that the greater the amount of shared visual information, the greater the comprehension of collaborators. However, it is important to note that many studies on video-mediated communication rejected this assumption (see section 3.2.2). Moreover, Karsenty (1999) provided empirical evidence of the cooperative nature of the communication process. He showed how speakers in a dialogue adapt the linguistic content of their speech to the hearer's cognitive environment. In this *cooperative view* of communication, the ease with which help requests are understood does not uniquely depend on the amount of shared visual information, but to the available shared resources which include non-visual contextual clues. Karsenty concluded that human-human dialogue is *doubly adaptive* (in the production and in the interpretation phase): if the medium does not offer visual channels, the producer of a message will voluntarily add extra information to the message to reduce possible misunderstandings, and similarly the receiver will use information available for other channels to disambiguate messages. Therefore, a maximally shared visual environment is not required to obtain the best comprehension efficiency. As Karsenty explains (1999, p. 310):

> One conclusion that should be drawn from this study is that the challenge in designing computer-mediated communication systems is to identify the minimum communication channels necessary to provide remote collaborators with an optimal shared visual context (and not a maximal shared visual context, which was the assumption underlying the shared visual information view). ... In particular, this study suggests that a video link could be superfluous, as far as comprehension efficiency is concerned. On the other hand, a screen-sharing system should be particularly useful, especially when the novices' calls require experts to understand not only how the novices' problems occurred but also which state the novices' systems are currently in.

Another class of design concerns for shared visual spaces is related to *embodiment*: the way in which human cognition arises from the body's interaction with the world. Robertson (1997) recognized the importance of the body as the essential basis of all human action and interaction. The author argued that the defining constraint of technology supporting collaborative work at a distance should be the essential corporeality of human cognition. She defined *Embodied Actions* as classes of cognitive practices that are publicly and simultaneously available to the perception of the actor and others in a shared physical space. While performed in co-presence, these actions are perceptually reciprocal: they are both perceiving and perceived. This is not a given in most virtual spaces. As Robertson explained (1997, p. 217):

> In shared physical space we can predict how our actions are perceived by others because we can perceive them ourselves as we live them. In technology-mediated communication individual participants will always be acting in their local physical space at the same time as they act in virtual space. Self-perception, then, will require not just the assumed resources of the local physical space but the development of perceptual skills and the provision of perceptual resources to enable each individual to perceive their own actions as they appear to other participants. Put another way, a basic principle in the design of CSCW technology to support cooperative work over distance is that the perception by others of any individual's actions needs to be explicitly regarded as part of the same process, or act of perception, as that individual's perception of their own actions.

As Heath et al. clearly explained (2001), expanding media spaces to include features of the remote participants' environments did not provide satisfactory support for 'object-focused' collaboration, as participants encountered difficulties in making sense of each other's conduct even with seemingly simple actions such as pointing to objects. The same can be said for Collaborative Virtual Environments, where individuals could not easily determine what a participant was referring to. The problem arose from the difficulty in re-connecting an image of the other with the image of the object to which they were referring (Hindmarsh et al., 1998).

Providing a shared visual space can assist in remote collaboration, however many questions are still open, as the degree of 'sharedness' of remote environments seems to be strongly linked to the task that the collaborators have to perform. Also, the interaction mechanisms related to the sharing of the workspace, or its subclass of elements, can easily fragment the body conduct of its users with the result of endangering collaboration. These are the main reasons that led some researchers to concentrate on more effective ways to enable gestures at a distance.

3.4 Remote gestures

In the early 1990s, Tang conducted several observations of teams collaborating on the design of a human-machine interface while using a large notepad or a whiteboard (J. C. Tang, 1989). Using ethnographical observations focusing on the interaction of people and artifacts, he noticed several key processes in face-to-face design activity:

a) collaborators use hand gestures in a significantly complex system which allows them to encode and convey a variety of different types of information;

b) the process of drawing images is often more important than the result, and conveys meaning in its' very act;

c) the drawing space itself, becomes a tool for the mediation of communication and collaboration processes within the group;

d) there are a variety of concurrent, different activities that take place within the drawing space; and

e) the literal spatial layout of the drawing space in relation to the collaborators has a role in structuring their activity.

In my own observations of developers using sketches to collaborate, I observed consistently that engineers used pen strokes and hand gestures to represent various kinds of relations between components of the system they were working on. My conclusion: what was important was the process of sketching more than its end product (Cherubini et al., 2007).

Bekker et al. extended Tang's seminal work analysing the role of gestures during design meetings and to inform the design of groupware system that could support these activities (Bekker et al., 1995). They refined a coding scheme developed by McNeill (1992), assigning the observed gestures to four categories: **Kinetic** (the movements reproduces an action performance), **Spatial** (the movement indicates distance or location or size), **Point** (fingers point to some person, to some object or place), and **Other** (all other gestures not coded with the previous categories). They recognized that gestures do not always appear isolated but often as a sequence. They noted especially four sequences: **the Walkthrough** (a series of *kinetic* gestures used to describe the interaction between a user and a product); **the List** (a series of *pointing* gestures); **the Contrast** (a *pointing* to one hand and then the other) and finally **the Emphasis** (the use of *other* gestures to place the accent to an utterance being uttered). Bekker and colleagues (1995) found that gesturing occurred systematically during co-located design meetings and that pointing was one of the most frequent types of gestures produced during the interaction. They characterize these gestures as follows (Bekker et al., 1995, p. 162):

> Point gestures were often used to refer to objects, persons, places or ideas. These gestures were used when design ideas were discussed, and also when the team discussed meeting management issues. Point gestures to objects often referred to (parts of) documents (see Figure 3-7), parts of the whiteboard or information on the computer screen. In some cases the gestures referred to very specific part of a document, e.g., a word or sentence, whereas in other cases they referred to a more vague piece of information, e.g., some concept described in a document or an area on the whiteboard.

See figure 3-7 for an example of a vignette of a pointing gesture combined with an artifact under discussion. In the same study, Bekker and colleagues also defined a series of purposes for which these gestures might be actually used and what particular category was more frequently observed for each of these purposes. Pointing was the type of gesture used in the largest number of circumstances.

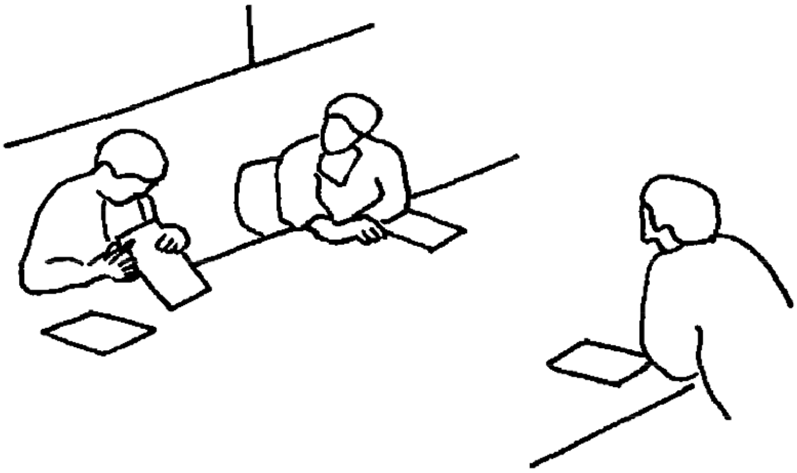


Figure 3-7: Pointing gesture combined with an artifact under discussion during a design meeting (from Bekker et al., 1995)

These studies highlight the importance of gesturing in collaborative work. Software aimed at sustaining collaborative work at a distance must adequately support the transmission of gestural information. This basic idea influenced the work of many researchers that tried to design systems to support remote gesturing using different mechanisms. One group of scholars looked at how gestures could be communicated at distance with the transmission of the collaborators' video-capture of the hands (see section 3.4.1). Other researchers tried to support remote gestures through digital metaphors (see section 3.4.2). Their basic ideas was that transferring free-hand drawings or showing digital artifacts representing the focus of attention to the partner at a distance could suffice to enable gesturing mechanisms.

3.4.1 Supporting remote gestures through video

In the early 1990s, there was tremendous momentum in the development of prototypes for supporting collaborative work at distance. Many designers employed the idea of using a video-feed to enable remote gesturing. The common denominator of these projects was the capture of the hands of the users and the fusion of the resulting video feeds. One of the earliest systems employing this concept was developed by Tang and Minneman and it was called VideoDraw (1991a). During the preparatory observations of design teams, the authors realized that much of the collaborators' activities involve hand gestures and that these gestures relate a sequence of events, or refer to a locus of attention, or mediate interaction. They stressed out that the importance of the produced marks decreases with time. Also, they noticed how these gestures were often conducted in relation to a sketch or to an object in the drawing space. Their idea was therefore to support hand sketches at a distance preserving the relationship between the gestures and their reference. VideoDraw followed this principle, allowing two people to share a drawing surface. It consisted of two drawing stations, each employing a video camera that captured the marks produced on the surface of a display. The digitized marks were displayed on the display of the complementary remote installation (see figure 3-8). As each collaborator drew on the screen, the video camera transmitted these marks and the accompanying hand gestures to the other collaborator. The prototype allowed concurrent access to the same part of the shared drawing. However, several limitations in using VideoDraw were observed. First, the prototype did not allow for the sharing of a great number of marks as the resolution was low. Second, the system allowed each participant to edit only her own marks. Finally, the tested prototype presented some rather uncomfortable human factors like the thickness of the glass, which added a consistent parallax distortion between the sketch on the surface and the retro-projected image.

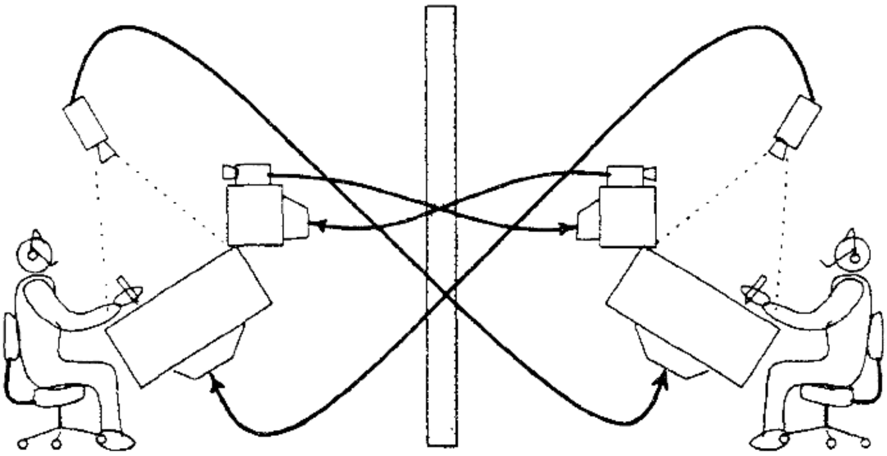


Figure 3-8: Schematic of VideoDraw system between two sites (from J. C. Tang & Minneman, 1991a)

Similarly, the VideoWhiteboard (J. C. Tang & Minneman, 1991b) system designed on the idea of the shadow plays of ancient China: it consisted of rear-projected screens that acted also as whiteboards. Users drew on the front of these screens and the marks and the accompanying hand gestures were imaged by cameras located on the other side. The captured video was displayed on the screen of the other participant (see figure 3-9). Although the idea was interesting, this system was not immune to the limitations of the video-based communication systems discussed above. Essentially, the authors noticed a fundamental asymmetry between the performed actions and the perceived actions. Pointing to detailed parts of the drawing was not supported by an appropriate resolution of the video cameras. Finally, gestures performed some distance away from the screen were difficult to perceive.

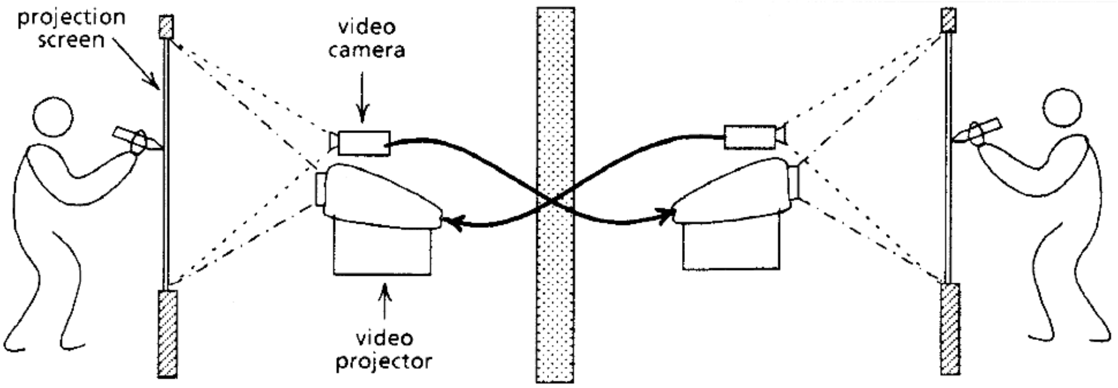


Figure 3-9: Schematic of VideoWhiteboard system between two sites (from J. C. Tang & Minneman, 1991b)

In the same years, Ishii and Miyake worked on a desktop-based system with a shared visual space for hand gesturing within a PC environment aiming at achieving the same flexibility of an actual workspace where collaborators can see each other as well as access objects and elements in the environment (Ishii & Miyake, 1991) (see figure 3-10). The system was evaluated through an experimental design that revealed a positive effect in supporting collaborative work. Particularly, users commented positively on the way the system supported reference resolution.

A year later, Ishii et al. worked on a refined system called ClearBoard, that allowed users to collaboratively sketch on a shared display while maintaining eye-contact (Ishii & Kobayashi, 1992; Ishii et al., 1993). The prototype was a response to the importance of eye-contact to interaction regulation, as the researchers explained (p. 526):

> Lack of eye contact has been another problem of existing desktop video conference systems. People feel it difficult to communicate when they cannot tell if the partner is looking at him or her. Eye contact plays an important role in face-to-face conversations because "eyes are as eloquent as the tongue."

Eye-contact allows the users to switch their focus smoothly from one to the other according to the task content. The ClearBoard system was designed as a transparent display replicating features of face-to-face interaction, and allowed the users to draw at the same time and to indicate on points of the drawing(see figure 3-11). One of the limitations of the systems discussed above was the *display disparity*, an unequal access to the shared sketch. Users of these systems were able to modify content that they themselves had created. One of the latest reiterations of ClearBoard introduced a groupware drawing tool in conjunction with a stylus. This allowed users to have equal access

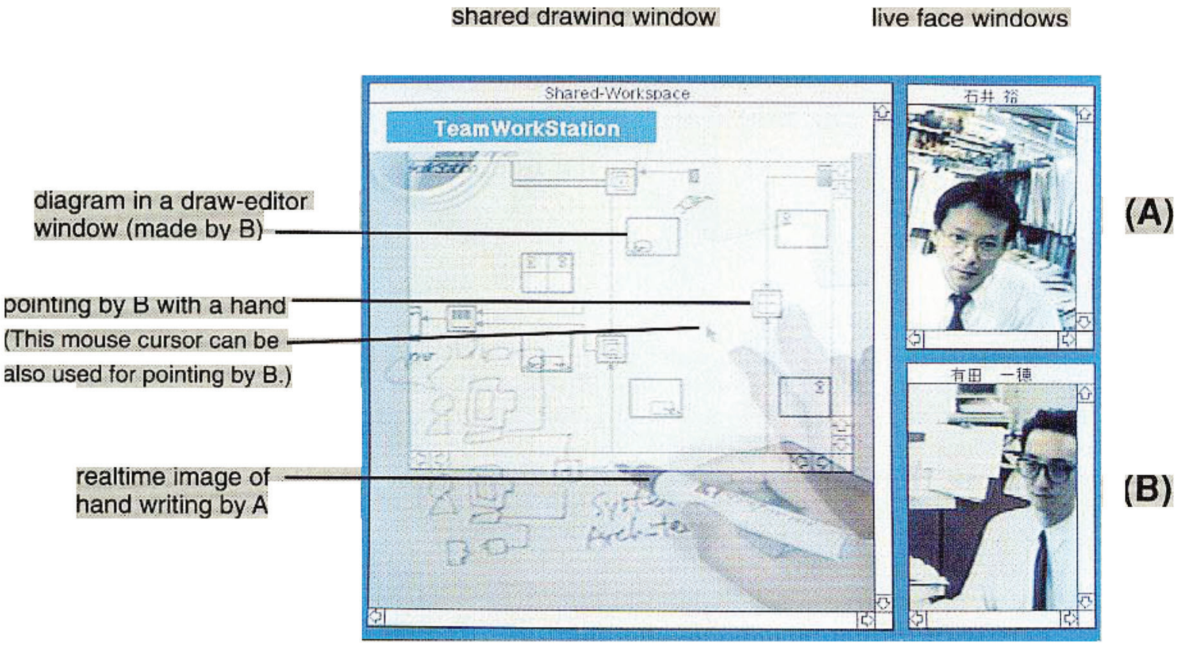

shared drawing window
live face windows
Shared-Workspace
TeamWorkStation
石井 裕
有田 一穂
diagram in a draw-editor window (made by B)
pointing by B with a hand
(This mouse cursor can be also used for pointing by B.)
realtime image of hand writing by A
(A)
(B)

Figure 3-10: TeamWorkStation shared screen in design session (from Ishii & Miyake, 1991)

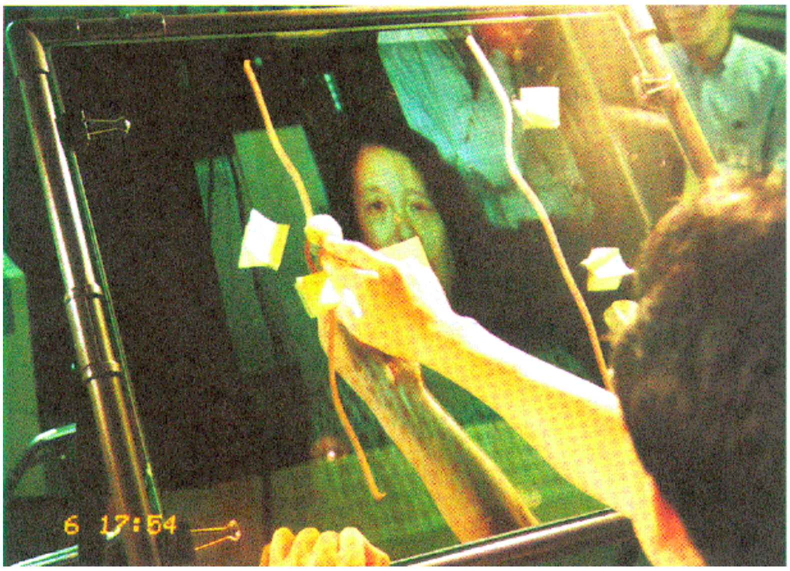


Figure 3-11: ClearBoard in use (from Ishii & Kobayashi, 1992). The key point of this system was to keep *seamlessly* in the same workspace the participants gaze, the gestures and the manipulation of the task artifacts

to the drawing. Each participant could control every component of the sketch and edit it. The system resolved the limitations raised by previous research. Bekker and colleagues noticed, for instance, how ClearBoard could potentially yield better results than face-to-face interaction because it created a seamless arrangement between participants and the shared workspace (Bekker et al., 1995, p. 165). Still, the system had some flaws as it could not scale easily beyond two users, and the expensive workstation implemented proprietary practices of use and protocols for interaction.

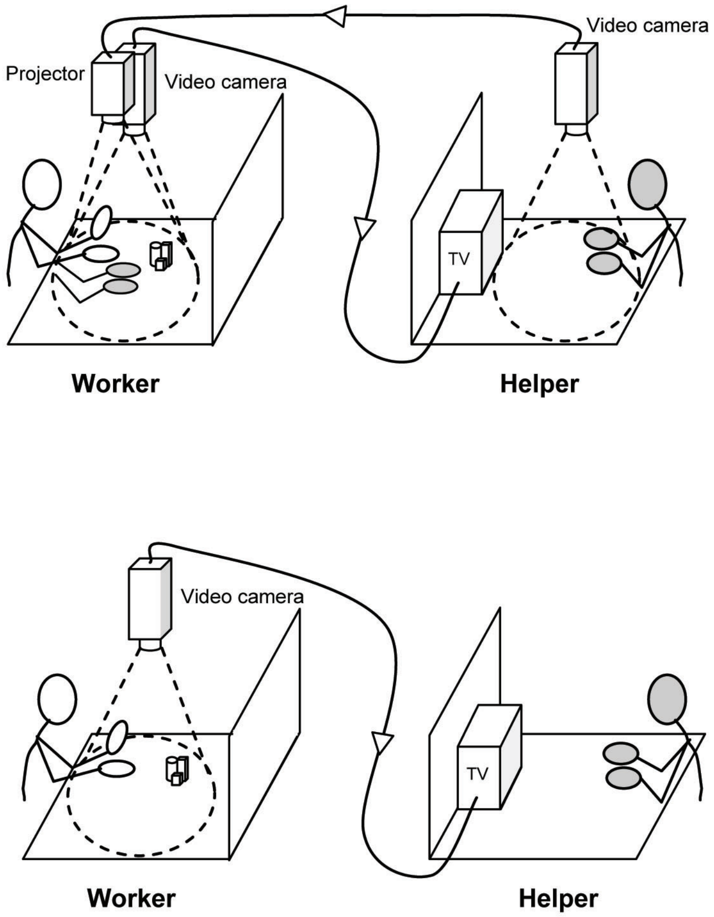


Figure 3-12: Experiment setup used by Kirk et al. (2007): above, voice plus projected hands condition; below, voice only communication condition (helper retains visual access to workspace)

More recently, Kirk and colleagues researched the ability of remote gesturing tools to improve distance collaboration performance (Kirk, 2006; Kirk et al., 2007). The authors built their work on the argument of Fussell et al. stating that complex gestures rather than simple deixis are responsible for performance enhancement of remote collaboration (Fussell et al., 2004). However, it must be noted that Fussell et al. derived this argument from the lack of performance enhancement in a controlled experiment where a *telepointer* was used to enable the remote gesturing in an experimental condition (the mouse pointer on a local machine is transferred to the screen of the remote display). They hypothetized that the part of the task requiring effective pointing represented only a small percent of the total and therefore was not influential. Their argument was purely speculative as they did not provide any experimental evidence that this finding could be generalized to other domains or tasks. To come back to the work of Kirk et al., their specific question was to understand the impact of complex remote gestures on language taking into consideration the temporal nature of the grounding process. Particularly, the authors argued that performance benefits that can derived from the use of a remote gesture tool are due to its ability to affect the process of developing common ground (Kirk & Fraser, 2006). Complex use of gestures in interaction can have a variety of other uses in collaborative discourse such as helping to marshal turn-taking and to signal understanding. The authors used a helper-worker paradigm, similar to that used by Gergle (2006), where the task at hand was the reconstruction of a Lego © model from diagrammatic instructions (see figure 3-12).

As they explain (Kirk et al., 2007, p. 1042):

> The system was constructed such that both participants would be in the same room during the study, but only had visual access to each other and each other's desks through the mediating technology – partitions ensuring that direct visual access was blocked. This enabled us to retain full audio in all conditions without having to use any audio communications technology. Participants were allowed to speak to one another at all times during the study.

They demonstrated that the performance benefits of remote gesture tools appeared to be strongest during early stages of an interaction, when remote gestures have the potential to reduce the amount of workers' speech. Independent of the phase, questioning behavior from the workers was slightly lessened by gesturing. Also, gesturing was associated with a reduction in the occurrence of speech overlaps. Their findings demonstrated that performance improvements, as already demonstrated by Fussell et al. (2004), still occurred when the remote gestures format was altered from a digital sketch to an unmediated representation of hands.

Another interesting line of research has been carried out by Kuzuoka and colleagues over the last decade. In 1992, Kuzuoka presented an investigation of the SharedView system that was developed to support remote collaboration in a three dimensional space (Kuzuoka, 1992). Kuzuoka demonstrated that remote gesturing, enabled by this system, could lead to higher performance and that the process of collaboration could be influenced by the availability of remote deictic technologies. His solution to support remote gestures was similar to that of VideoDraw discussed above but also SharedView additionally employed an head-mounted camera on the operator's side, which allowed for a focal point and the minimizing differences in directional expressions (see figure 3-13).

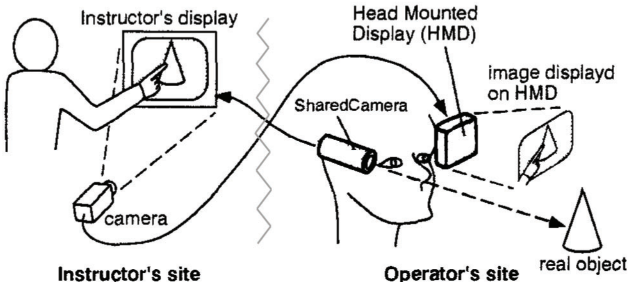


Figure 3-13: SharedView system (Kuzuoka et al., 1994). The operator wears the SharedView. The SharedCamera's image is sent to the display at the instructor's site. The instructor uses gestures in from of the display, which are imaged by a camera and sent back to the operator's HMD. In this way, the instructor can give instruction with gestures

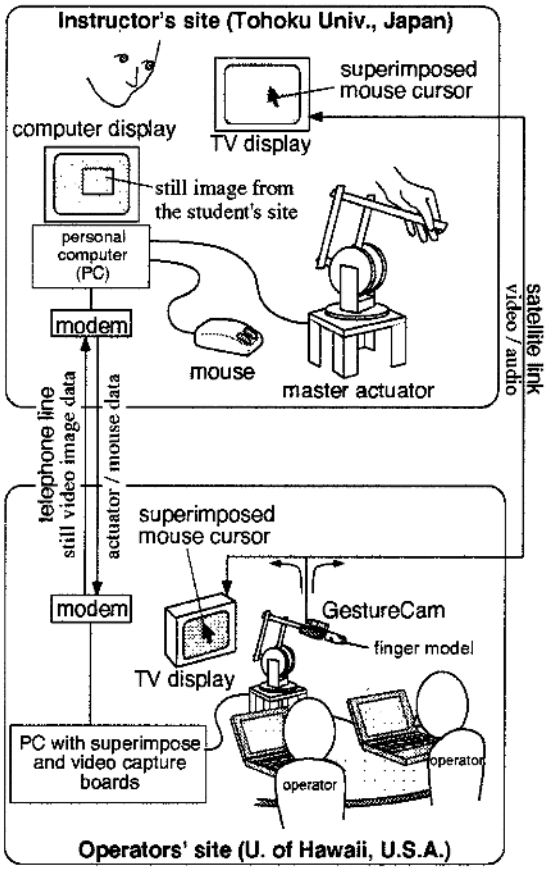


Figure 3-14: GestureCam system (Kuzuoka et al., 1994). Setting for distance instruction experiment

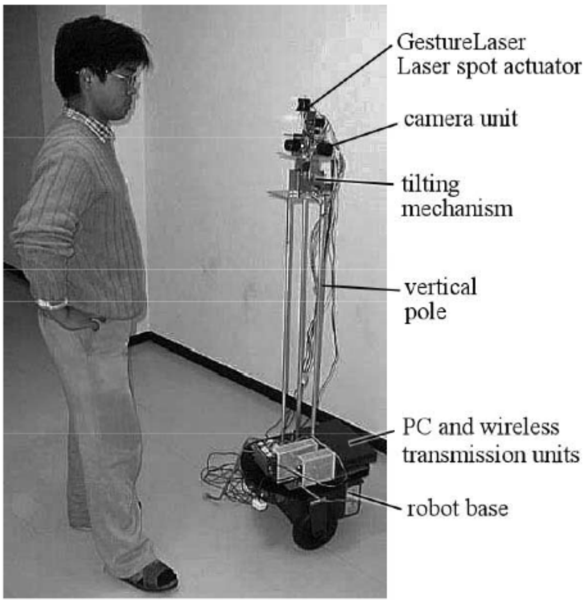


Figure 3-15: GestureMan system (Kuzuoka et al., 2000)

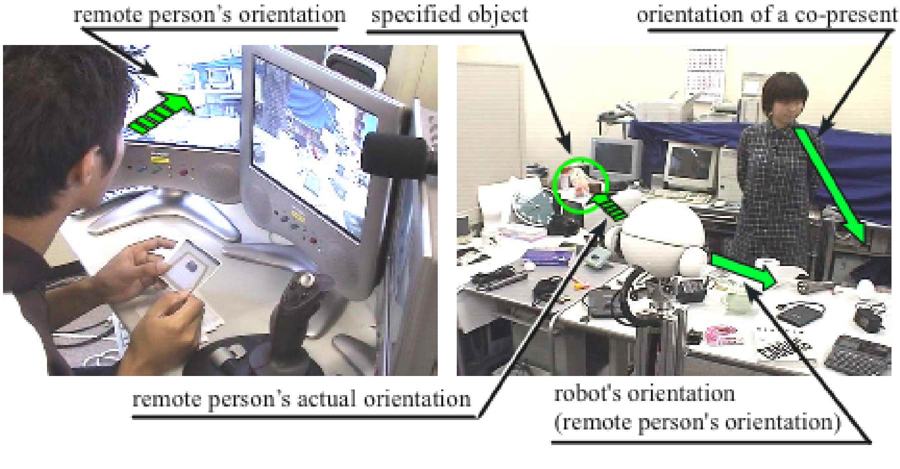

remote person's orientation
specified object
orientation of a co-present
remote person's actual orientation
robot's orientation
(remote person's orientation)

Figure 3-16: An example of false anticipatory reaction observed in the use of the GestureMan system (Kuzuoka et al., 2004). The robot's head has three cameras, therefore the remote instructor can see the target object, while the local operator relies on the direction of the robot's head to locate the target object thus resulting in a mismatch

Although these features were extensively described as being of benefit for remote instruction, Kuzuoka et al. critiqued the system in a later work (Kuzuoka & Shoki, 1994). Users complained that the system offered a too narrow view of the collaborative workplace. The instructors could only see what the operators wished them to see. To counter these problems, they designed a new system, called GestureCam (Kuzuoka et al., 1994). In this system, the instructor could remotely control the direction of an otherwise static camera in the operator's workplace. Attached to this camera, there was a laser pointer and a finger that represented the action of pointing at a given place (see figure 3-14). The use of the system was hindered by a poor video-link between the remote sites. However, an evaluation of the system in use revealed benefits of the embodiment the instructor's point of view through the robotic arm, a theme that led to further research (Kuzuoka et al., 1995).

In further work, Yamazaki et al. (1999) refined the GestureCam system by focusing more on the utility of the laser pointer. The GestureLaser system made this pointer independent of the camera view. The authors' intent was to allowing for an increased range of viewing and gesturing functions for the user. Through their research, this group of scholars became progressively more interested in the embodiment of gestural actions in remote space. This led to the development of the GestureMan system, where the laser pointer and the camera were embedded in a robot that could be controlled by the remote operator (Kuzuoka et al., 2000) (see figure 3-15). Ironically, the analysis of use of this system demonstrated that users' conduct became disembodied, and therefore problematic (Kuzuoka et al., 2004). Orientation and reference to the task artifacts was problematic (see figure 3-16).

3.4.2 Supporting remote gestures through digital metaphors

In parallel to the development of solutions aiming at enabling remote gestures through video, other researchers become interested in the idea that gesturing could be efficiently supported by digital sketches or metaphorical representations of the hands. One of these early systems was Commune, a system allowed users to sketch on a common drawing surface while talking (Bly & Minneman, 1990). Strokes produced with a stylus were digitalized and visualized on the screen at the remote location (see figure 3-17). The authors carefully observed the use of the system by several pairs and they found that neither talking nor the marks alone effectively communicated the issue under consideration, but that users could seamlessly integrate their speech to the activities on the workspace (see figure 3-18). Interestingly, the system had an advantage over a sketch realized during a face-to-face meeting, because the users could draw over the same portion of the sketch at the same time. While this is possible with the virtual superimposition of the marks, the same is not possible in real situations as the collaborators' hands are in the way. The authors speculated that this could have been an advantage over co-located meetings. However, they also identified three problems of the system: the styli were difficult to use with sketches consisting of short line segments; the drawing space was small for certain activities, and finally Commune constrained gestures to pointing actions, because collaborators could only see marks over the drawing surface.

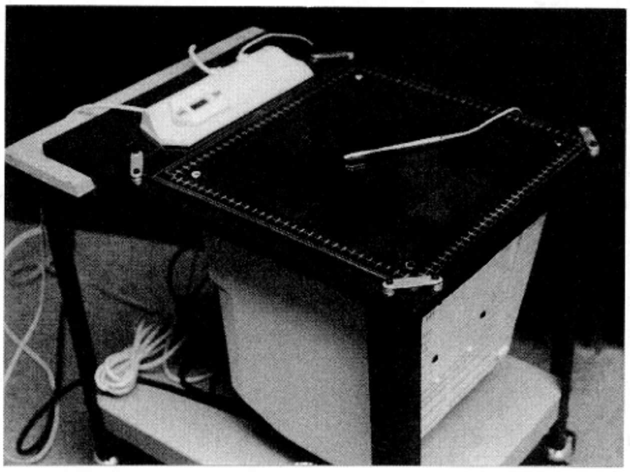
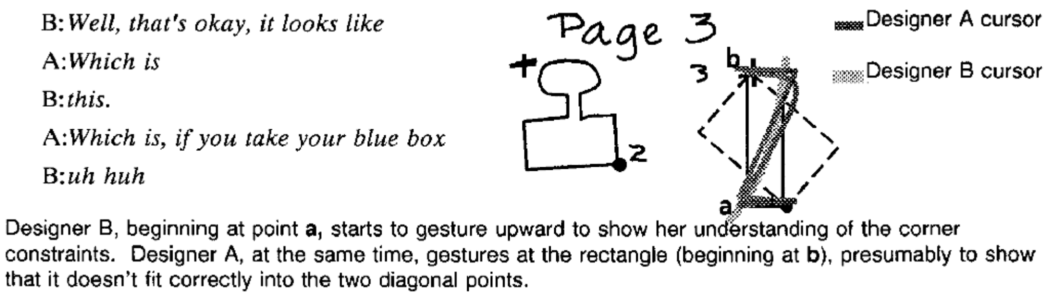


Figure 3-17: The Commune workstation consists of a horizontally-oriented monitor with a digitizer (Bly & Minneman, 1990)



B: *Well, that's okay, it looks like*
A: *Which is*
B: *this.*
A: *Which is, if you take your blue box*
B: *uh huh*

Designer B, beginning at point **a**, starts to gesture upward to show her understanding of the corner constraints. Designer A, at the same time, gestures at the rectangle (beginning at **b**), presumably to show that it doesn't fit correctly into the two diagonal points.

Figure 3-18: Commune users closely intertwined talk and drawing in surface use

While the prototypes described above considered remote gestures as conversational resources that could be used by collaborators in *specific* moments of the interaction, other researches advanced the idea that gestures in the workspace could be defined, and therefore supported, as a generic knowledge of *what* the others were doing regardless of specific communication events. The ensemble of these perceptual cues that can help collaborators to keep track of what the others are doing goes under the name of *workspace awareness*, as described in section 3.3.2.

Gutwin and Greenberg published a number of researches where they reported mixed effects of supporting workspace awareness for collaborative work at a distance (Greenberg et al., 1996; Gutwin & Greenberg, 1999). One of their solutions consisted in showing in each collaborator's interface a minimap of the interface of the other. This visualization reported in a schematic way the basic elements of the interface plus the information of where the other was looking. A semi-transparent rectangle (*viewport*) represented the elements of the workspace that were actually visualized in the collaborator's screen and a mouse pointer (*telepointer*) showed the elements of that part of the workspace the collaborator was interacting with. Figure 3-19, taken from (Gutwin & Greenberg, 2004), represents the *minimap* used by Gutwin and Greenberg. The authors specifically highlighted how workspace awareness is related to deixis and gaze and affects the conversational common ground necessary for the interaction (Gutwin & Greenberg, 2004, p. 12):

> The role of workspace awareness in deixis (i.e., where one's pointing or gesturing action disambiguates conversational references, such as when one says "this one" while pointing to an object), visual evidence and gaze awareness means that the elements of awareness are part of conversational common ground in shared spaces (Clark, 1996). This implies that not only do you have to be aware of me to interpret my visual communication, but that I have know what you are aware of as well, so that I can safely make use of the workspace in my communication.

From the experimental point of view, a more systematic study of the impact of telepointers on learning at a distance was recently carried out by Adams et al. (2005), who demonstrate that the presence of a telepointer in an experimental course improved the students' recall of the contents taught (the teacher was not co-located with the students and used the telepointer to highlights parts of sketches presented through a slideshow).

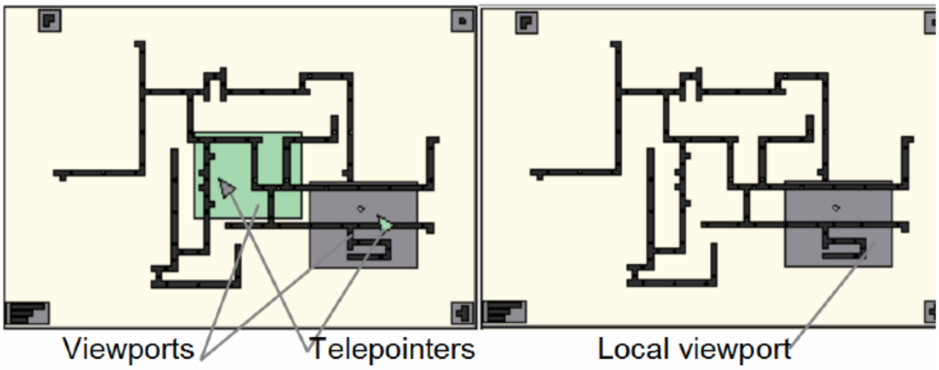


Figure 3-19: Minimap used by Gutwin and Greenberg. Radar view (left) and Overview (right)

However, others evaluations of the use of the telepointer to support collaborative work at a distance reported negative results. For instance, Tang et al. developed a prototype called MPGSketch that allows mixed presence groups (e.g., some people working in presence with other teammates at a distance) to create and share sketches (A. Tang et al., 2004). Its aim was to remove issues of previous research like *display disparity* (unequal access to the functions of a shared workspace).

An evaluation of the system revealed weakness of the telepointer mechanism as a gesturing device, because that cursor activity could not be always related to the user intention, attention or presence. Essentially, the tool presented some embodiment issues that lead to users' access and control disparities. A second iteration of the technology, renamed Digital Arm Shadows, combined the digital sketch features of the previous prototype with the capture and transmission of body movements. They found that this second prototype was effecting in signalling participants' presence.

Following the same line of research, Fussell et al. (2004) recognized the importance of gestures in remote collaboration and distinguished between pointing gestures and representational gestures. This second group could be further divided into iconic representation, spatial gestures and kinetic gestures. The authors chose to adopt a surrogate approach to remote gestures, expressing the communicative intents through sketches rendered electronically through a digitalizing surface and a stylus. The reason for this choice over a more natural representation of gestures was supported by evidence that such alternative representations incorporate visible *embodiments* of gesture, and are therefore as good as higher-fidelity representations. In a preliminary study, the authors compared a face-to-face interaction to a remote collaboration where the helper could indicate parts of a model using a telepointer mechanism. Contrary to their hypothesis, they found that adding the pointer was not sufficient to improve performance over that of the video-only condition. They discussed possible explanations for this, like the fact that the cursor tool was too limited in functionality. Also they hypothesized that the part of the task requiring effective pointing represented only a small percent of the total and therefore was not influential.

These findings led the authors to develop the Drawing Over Video Environment (DOVE) that was presented and discussed in Ou et al. (2003). The workspace in the DOVE system is captured by an IP camera. This video feed is then displayed on a Helper's tablet PC, which is used to write or draw over the images with a digital pen. The resulting annotated images are displayed on a monitor located in the Worker's space (see figure 3-20). In the second experiment reported in (Fussell et al., 2004), they evaluated the DOVE system. They compared video-only against the DOVE setup with manual erasure and with automatic erasure. They found automatic erasure resulted in the best performance. Further analysis demonstrated that the majority of the actions performed by the Helpers during the task were pointing gestures (see figure 3-21).

One last note on this particular category of application concerns the work of Wellner (1993). His idea was that instead of making a personal computing environment look more like a desk, the opposite was also possible: giving an actual table top the computing power of a personal computer. In the early 1990s, he designed many conceptual prototypes for what he called the Digital Desk. One of the strengths of this particular approach was the ability of using embodied actions, like pointing gestures as an interaction mechanism (see figure 3-22). Fifteens years later,

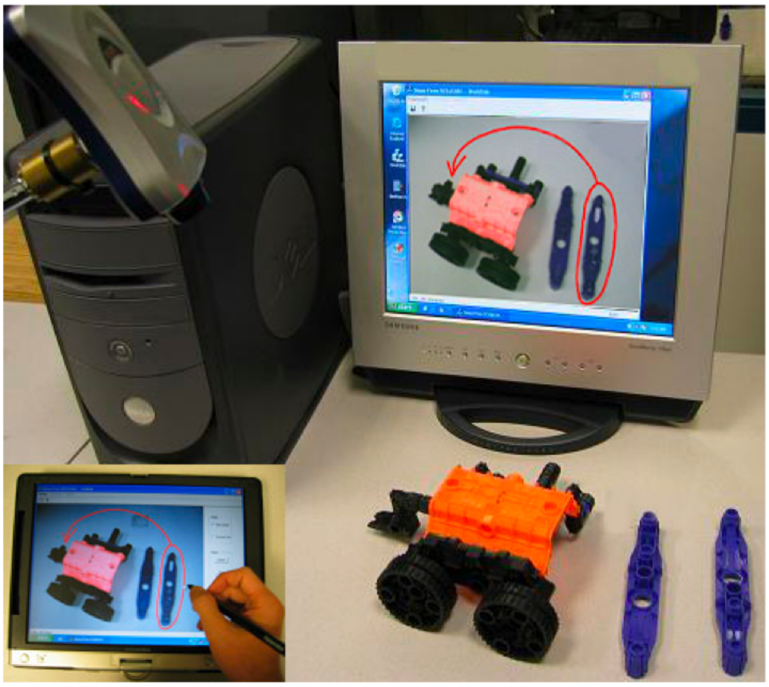


Figure 3-20: Close-up of the DOVE drawing tool on the Helper's tablet PC (left front insert) and on the Worker's monitor (right) (from Ou, Chen, et al., 2003)

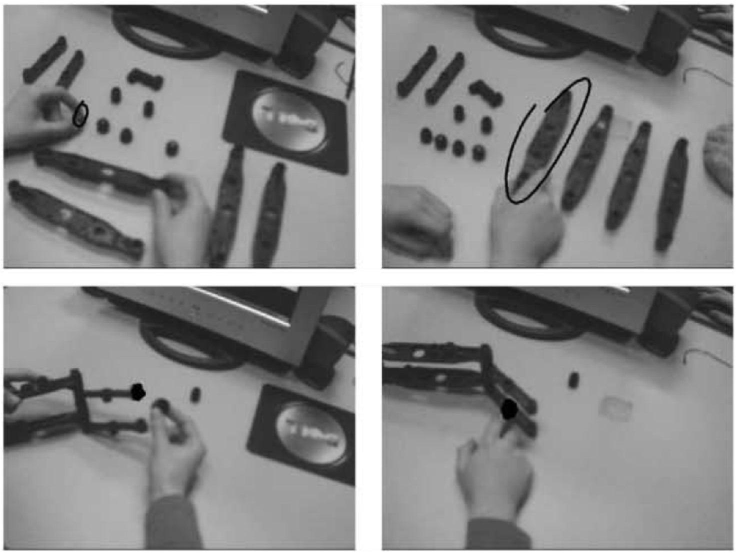


Figure 3-21: Examples of pointing annotations in the DOVE environment (from Fussell et al., 2004). Pointing to task objects (above), and pointing to target locations (below)

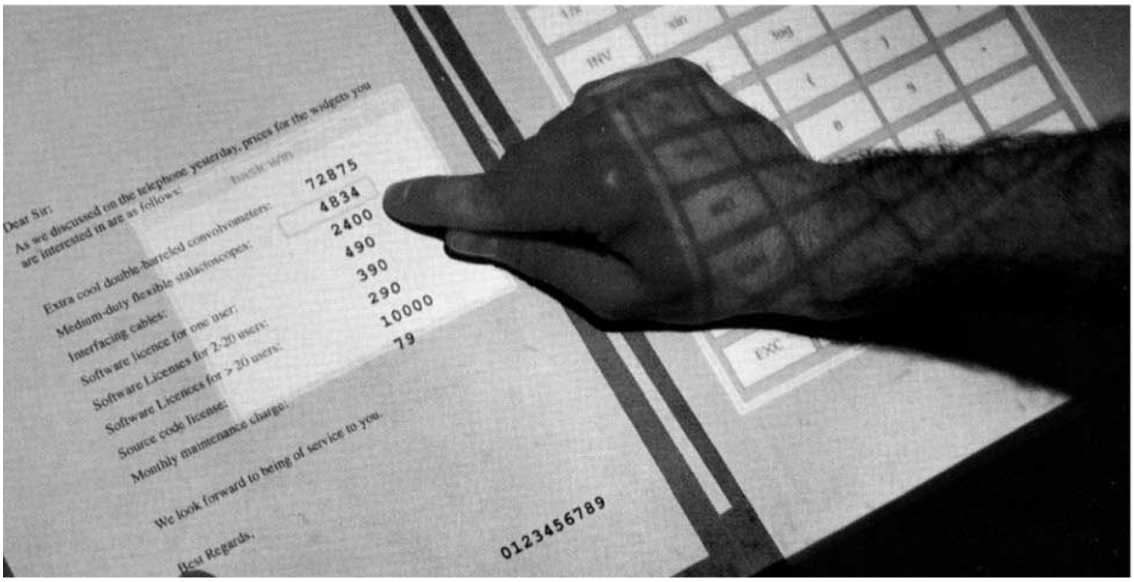

Dear Sir:
As we discussed on the telephone yesterday, prices for the widgets you are interested in are as follows:
Extra cool double-barreled convolvometers:
Medium-duty flexible stalactoscopes:
Interfacing cables:
Software licence for one user:
Software Licenses for 2-20 users:
Software Licences for > 20 users:
Source code license:
Monthly maintenance charge:
We look forward to being of service to you.
Best Regards,
72875
4834
2400
490
390
290
10000
79
0123456789
INV
sin
log
EXC

Figure 3-22: Pointing gesture on the Digital Desk. The number 4834 was selected by the user (from Wellner, 1993)

Apple Inc. designed and commercialized tactile interfaces which use the same design principle[1].

3.4.3 Synthesis: design issues of remote gesture tools

Gestures represent an extremely important mechanism that allow people to coordinate their efforts and disambiguate their contributions. Indeed, the research presented in this section presents mostly positive results on the use of technological means to enable gestures at a distance to sustain collaborative work. In the last twenty years, many solutions have been proposed to support remote gestures. Many of these early prototypes used video technology to capture and display the hands of the collaborators at the remote sites. Because of this choice, many of the issues that we discussed before stayed unsolved and undermined the large-scale adoption of these prototypes. Also, new human factors arose that will require further research (e.g., the virtually overlapping hands allowing concurrent access to the same part of the drawing, which not usually possible when face-to-face). As Luff et al. (2003) clearly explained, video solutions suffered from a fracture of the ecology of the remote sites. In such systems, conduct was *fractured* from the place where it was produced and where it was received. Restricted field of views and distortion of projection are just few examples of how video can hamper the usefulness of remote gestures.

Other researchers envisioned prototypes using digital metaphors of gestures like digitized sketches or pointers. The underlying hypothesis of these studies was that a sketch could incorporate features of a gesture that could suffice to replace the real gesture. Although results in

[1] I am referring to the interface of the iPhone, which enable multi-finger interaction to control applications.

this direction are promising, this hypothesis is still unverified in the literature. More research is required to define what features of gesture are most important for communication, what kinds of gestures are necessary and in what circumstances. Particularly as the issue of enabling gestures has often been tackled at global level, little attention has been given to the role of specific forms of gestures and their interplay with other facets of human interaction.

In this regard, deictic gestures have often been under-considered while striving for enabling the full range of gesturing features at a distance. However, the importance of deixis and deictic gestures for collaboration was observed and confirmed in many studies, as I will discuss below. Finally, one issue that it is still insufficiently regarded is that of the interconnection of gestures with language. Human always use a multiplicity of communication channels. Gestures are often, if not always, accompanied by speech and monitored through gaze. Therefore, more research is needed to understand these interconnections and how to best exploit them to sustain remote collaboration.

3.5 Remote deixis

In this section, I want to report on studies examining the role of remote deixis (in this thesis called Explicit Referencing[2]) on collaborative work at a distance. The relevant contributions can be organized into two categories: studies approaching the problem at a *linguistic level*, concentrating on few limited utterance exchanges, and those approaching the problem from the *collaboration level*, observing longer interactions during complex tasks. While both approaches contribute to the understanding of human cognition and interaction mechanisms, the difference in the scale of analysis often yields divergent results. While studies targeting the linguistic level focus on the dialogue interchange occurring between two or more interlocutors, studies targeting the collaboration level refer to the shared understanding that is constructed as a consequence of that exchange (Dillenbourg & Traum, 2006). I have discussed this issue in the previous chapter.

Another criteria that I used to structure the relevant works presented here is the organization mechanism used in interfaces to present users' contributions. As I will show in chapter 4, linguistic contributions linked to a geographical context can be organized *by time*, therefore following the temporal order of production, or *by space*, therefore following the physical area of interest to which similar messages are linked. The choice of the latter model can have consequences for the flow, or linearity of the conversation, while the choice of the former can split the user's attention between the workspace and the conversation, again with negative consequences.

[2] When an interface is designed to allow a specific user's message to be visually linked to a region or an artifact in the shared workspace, then I say that it implements Explicit Referencing. Explicit referencing is a general concept that is closely related to several notions such as artifact-centered discourse (Suthers & Xu, 2002), anchored discourse (Guzdial, 1997), anchored conversations (Churchill et al., 2000), or document-centered discourse (Buckingham-Shum & Sumner, 2001).

3.5.1 Explicit Referencing at linguistic level

People working together to solve a problem need a shared language to communicate. They also need to coordinate their activities, defining common goals and strategies to achieve them. Clark developed a theory describing how conversational partners develop a shared understanding, building shared knowledge or common ground (Clark, 1996). Clark defined the process of reaching this common ground, called grounding, as the effort of the conversational partners to share their attitudes, beliefs, expectations and mutual knowledge (Clark, 1996, Clark & Shaefer, 1989).

Clark and Brennan argue that the effort and the ease required to maintain a common ground throughout collaboration are critically dependent on the features of the media the conversation participants use to communicate (Clark & Brennan, 1991). For example, the media can influence the listener's ability to offer feedback or to provide or seek clarification. The degree of sharedness of a visual space or the possibility of making deictic gestures are features of the communication media that influence the grounding mechanisms. Let me consider a case where two peers are discussing where to meet by mobile phone. The first is guiding the second to a meeting point and is offering detailed information. The second is following this information to reach the first speaker. Without visual contact, the first speaker will tend to use a detailed description of the landmarks with a consequent high effort and nonetheless a high probability of misunderstanding. In a different situation, if the peers share a map over which the first speaker can use deictic gestures, the resulting dialog will be much lighter in terms of number of words used and effort required. All visible elements in the shared visual space become part of the visual information of the task. In a face-to-face collaboration, the shared visual space is composed and influenced by the artifacts used during the participants' interaction and the participants themselves. Their body movements, *proxemics*[3], gestures, facial expressions, and gaze all play a role in the establishment of the common ground.

In terms of the pointing gestures, Clark (2003) explained that communication is ordinarily anchored to the material world and that one way it gets anchored is through *pointing*. Clark also explains that there exists a counterpart of pointing: *placing*. Through the use of our position and the position of the objects we refer to in the actual world, we shape context to reduce misunderstanding and we make communication more efficient (e.g., placing shopping items on the counter). He argues that *directing-to* and *placing-for* are two indicative acts. Indicating has fundamentally to do with creating indexes for things. Every indication must establish an intrinsic connection between the signal and its object. The more transparent is this connection the more

[3] The term proxemics was introduced by anthropologist Edward T. Hall, in 1959, to describe set measurable distances between people as they interact. The effects of proxemics, according to Hall, can be summarized by the following loose rule: *"Like gravity, the influence of two bodies on each other is inversely proportional not only to the square of their distance but possibly even the cube of the distance between them"*.

effective is the act. That is why we cannot use an indication to an object without the originating signal. Finally, indicating an object in space must also lead the participants to focus attention on that object. In other words, anything that focuses the attention is an index.

This implies that effective indicating gestures should attract eye movements. However, gazing is not just a perception device. Clark and Krych (2004) highlighted how gazing is a communication device used to designate the person or things the speaker is attending to, or used to monitor the addressees' understanding while one is speaking. Also, eye gaze as a communicative act is not effective unless the person being gazed at registers it. So we often talk of mutual gaze. Similarly, Richardson et al. (2005) demonstrated how the eye movements of a listener following a speaker monologue were significantly related with the eye movements of the speaker over the same visual scene. They also demonstrated that the degree of this coupling was related to the listener's performance on comprehension questions.

Considering the sight of artifacts in the workspace, visual information has been described by Clark and Marshall (1978) as one of the strongest sources for verifying mutual knowledge. Visual information can also be used to coordinate the shared language with which objects and locations are described (Gergle, 2006). For example, if an utterance is ambiguous in a certain context (e.g., "take the red book on the table", with multiple reddish books on the same table) this can be easily disambiguated by joining a deictic gesture to the contribution (e.g., "take that book"), with a subsequent economy of sentence-production and grounding effort.

Communication media limits the visual information that can be shared, with resulting effects in the collaboration process and performance. To test this hypothesis, Kraut et al. (2000; 2003) conducted two experiments using a bike-repair task where an expert was guiding a novice repairing a bike under various communication configurations: audio-only, and a second condition where the 'Helper' could see a video taken from a camera mounted on the helmet of the 'Worker'. They had pairs side-by-side in the control condition. Communication was more efficient in the side-by-side condition, where the helper spent more time telling the worker what to do. In the mediated condition, not only were the dialogues longer, but their focus also shifted: more speaking turns were devoted to acknowledging the partners' messages. Their results indicated that physical tasks could be performed most efficiently when a helper is physically co-present. Having a remote helper leads to better performance than working alone, but having a remote helper is not as effective as having a helper working by one's side. The visual information was valuable for keeping the helper aware of the changing state of the task (see figure 3-23).

Gergle et al. (2004a) presented a study that demonstrated that action replaces explicit verbal instruction in a shared visual workspace. In their experiment, pairs of participants performed a referential communication task with and without a shared visual space. A performed a sequential analysis of the messages and actions of the different trials, and revealed that pairs with a shared

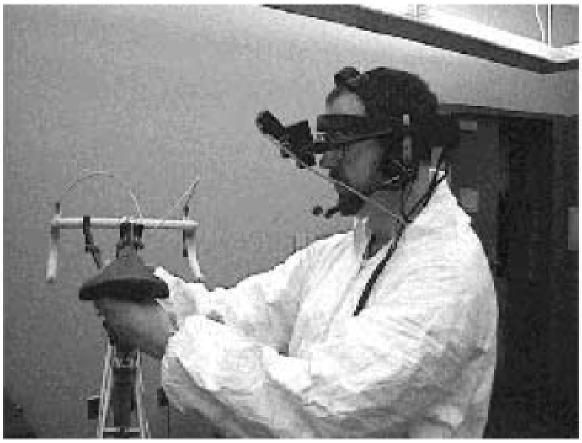


Figure 3-23: Experimental setup used by Kraut et al. (2003). Worker wearing the collaborative system

workspace were less likely to explicitly verify their actions with speech (e.g., provide and seek verbal acknowledgements from the collaborator). Instead, pairs that had access to a shared visual space relied on visual information to disambiguate references used to guide their partner.

Deictic gestures are naturally produced in the visual space shared between collaborators. These are always combined with messages, as they are used to disambiguate and enrich the linguistic content. Brennan (1990, 2004) devised an experimental task where two participants had to interact at distance, coordinating their actions over a shared map in order to park two icon-cars on the parking lot. She showed that the use of a telepointer increased the speed at which the remote collaborators could match the icons, but lowered the accuracy of the final result, since both users knew they were close to each other on their screens, while the non-telepointer pairs needed to be more explicit about each detail to be sure they were in the correct location.

When the interface used by the remote collaborator does not support deixis, collaborators often rely on communication strategies to explicit the references used in the interaction. Kraut et al. (2002), using their helper-worker puzzle task, found that the use of 'spatial-deixis-terms', phrases used to refer to an object by describing its position in relation to others, such as "next to", "below", or "in front of", was substantially higher in the absence of a shared visual space, since this was one of the primary ways in which the pairs could describe the layout.

3.5.2 Conversation linearity and turn taking at linguistic level

However, as briefly explained in the introduction of this section, different implementations of the mechanism to link the communication exchange to the shared context can result in a non-linear conversation with consequences on the task performances. In Computer-Mediated Communication linearity is an orthogonal dimension to the persistence of the conversation. Chat applications often display a certain number of previous messages, therefore supporting visually the temporal flow of the conversation and lengthening the persistance of the messages' content. Dialogue persistence has many effects on collaborative work: it reduces the cognitive load by providing an external representation of information jointly shared by the conversational partners (Dillenbourg & Traum, 2006); it provides a means for pairs to parallelize communication and actions (Gergle, Millen, et al., 2004), and it affects collaborative task performances (McCarthy & Monk, 1994).

The persistance of the conversation has also negative consequences for communication. Cherny (1999) explored the effects of real-time computer mediation on communication and the extent to which the MUD experience parallels face-to-face interaction. She studied turn taking in this medium and how participants cooperate to create a "floor" for conversation (the possibility of contributing to the conversation by explicitly recognizing the participants' attempts to take the floor). She showed that the absence of conventional face-to-face communication mechanisms make more equal level of participation possible, but also that the same medium makes it much easier to ignore other's contributions.

The way conversational partners alternate their turns of contribution was deeply studied in relation to the linearity of the conversation and to understand consequences for collaboration. Condon and Čech (2001) studied turn taking across different communication modalities (face-to-face environment, an asynchronous, e-mail environment, and several types of synchronous computer-mediated environments). They showed that when turns increase in duration participants switch from serial to parallel strategies to organize their decision-making. They also showed that pivot turns, which are turns that are much shorter than those that precede or follow them, can reflect the discourse functions of the relevant turns. Finally, turn duration can be used as a metric for measuring the dominance in conversation.

Similarly, Hancock and Dunham (2001) argued that communication settings that inhibit some turn-taking behaviors result in a loss of coordination between actors. Their experiment supports Clark's proposal that a communication setting that disrupts the regulation of turn-taking will both undermine higher level language processes (i.e., the construal of meaning) and increase the frequency of meta-communicative signals required to coordinate the speaker's action with the listener's attention.

3.5.3 Conversation linearity at collaboration level

While the works presented above targeted the understanding of the global role of the linearity of the conversation on communication, other research focused on the effects of the availability of this feature on collaborative work at distance. McCarthy and Monk (1994) presented a controlled experiment to assess the effect of dialogue history in a referential communication task. They found that a larger dialogue history enabled the pairs to reference utterances that occurred much earlier in the discussion.

Smith et al. (2000) presented a system to organize the contributions of a chat in a threaded interface. Instead of the classic time organization criterion used in the majority of chat applications, the authors experimented with a chat prototype that organized the contributions according to the topic that the users were discussing. They hypothesized that this tool would reduce the ambiguity of certain contributions due to intertwined turn taking. The results of their qualitative evaluation showed that patterns of interaction in threaded chat were equally effective, but different than standard chat programs. However, users rated their threaded chat as worse than a regular chat program.

Similarly, Fuks et al. (2006) developed a tool to avoid chat confusion (when it becomes difficult to follow the conversation as two or more topics become intertwined in turn-taking and result in an increase in the ambiguity of short answers), as the authors recognized in the irregular turn-taking a source of miscomprehension. The author proposed a chat tool, called Mediated Chat, where it is possible to regulate the turn taking using predefined modalities like ‘free contributions’, where each participant can send a message at any time; or ‘circular contributions’, where the participants are organized in a circular queue. The authors reported that chat confusion was more likely to occur during free contributions. More precisely, during the branching-out phase of the free conversation other topics were discussed that made confusion more likely to occur.

Additionally Phillips (2000) investigated the role of turn-taking formats in real-time text-only computer mediated communication with a particular focus on the tradeoff between smooth turn-taking exchanges and moment-by-moment collaboration between participants. He showed that, contrary to what popular models of dialogue would predict, users communicating with interfaces that imposed a turn-taking format produced less efficient dialogues and performed less well on collaborative brainstorming and recall tasks.

To conclude, while mixed results have been reported while trying to reduce chat confusion by organizing the messages according to their content or by restraining the turn-taking of the conversation, many scholars reported positive qualities of organising messages according to a temporal criterion and supporting the permanence of the conversation. Research in this area, demonstrated how collaborators adapt their conversations to restrained medium and also how these inhibitions of more spontaneous organization of conversation might have negative implications for collaboration. Figure 3-24 represents the two different criteria that I presented.

3.5.4 Explicit Referencing at collaboration level

Some studies have investigated the effects of referencing to the shared workspace on collaborative work at a distance: the more the objects the conversation refers to are visible and shared by the communication peers, the better the performance in the collaboration. Van der Pol et al. (2006b;

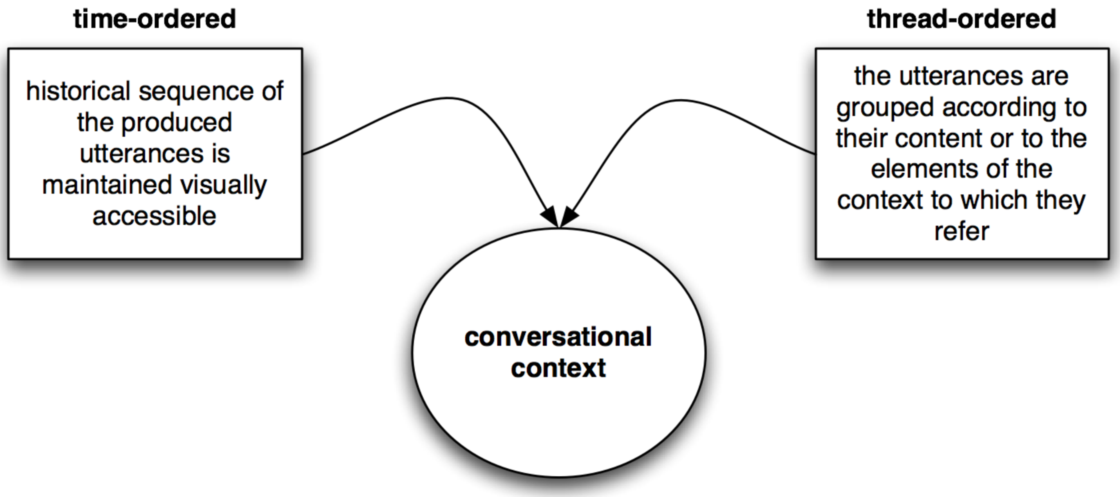
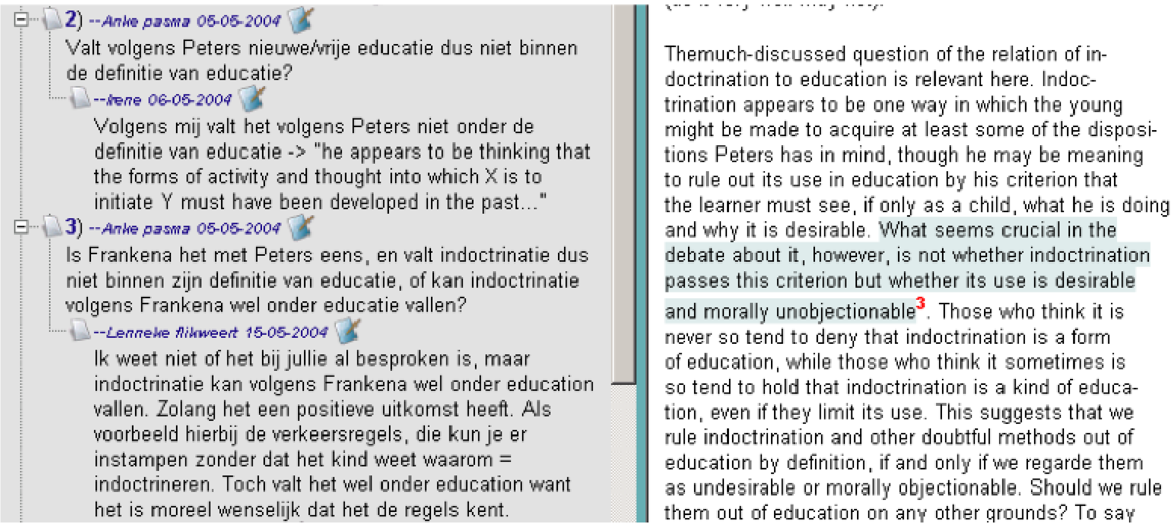

time-ordered
historical sequence of the produced utterances is maintained visually accessible
thread-ordered
the utterances are grouped according to their content or to the elements of the context to which they refer
conversational context

Figure 3-24: Organization of the messages of a conversation according to two different principles: emission time as opposed to content

2006a) researched context enhancement for co-intentionality and co-reference in asynchronous computer-mediated communication. The authors developed a tool for linking students' conversations to documents under discussion (see figure 3-25). Results indicated that the tool reinforced task-context by providing a frame of reference for the conversation and led to a smaller topic-drift in the answers posted to new topics in the forum. They concluded that for collaborative text comprehension, explicit referencing to task context is more suitable than traditional forum discussion.


2) --Anke pasma 05-05-2004

Valt volgens Peters nieuwe/vrije educatie dus niet binnen de definitie van educatie?

--Irene 06-05-2004

Volgens mij valt het volgens Peters niet onder de definitie van educatie -> "he appears to be thinking that the forms of activity and thought into which X is to initiate Y must have been developed in the past..."

3) --Anke pasma 05-05-2004

Is Frankena het met Peters eens, en valt indoctrinatie dus niet binnen zijn definitie van educatie, of kan indoctrinatie volgens Frankena wel onder educatie vallen?

--Lenneke flikweert 15-05-2004

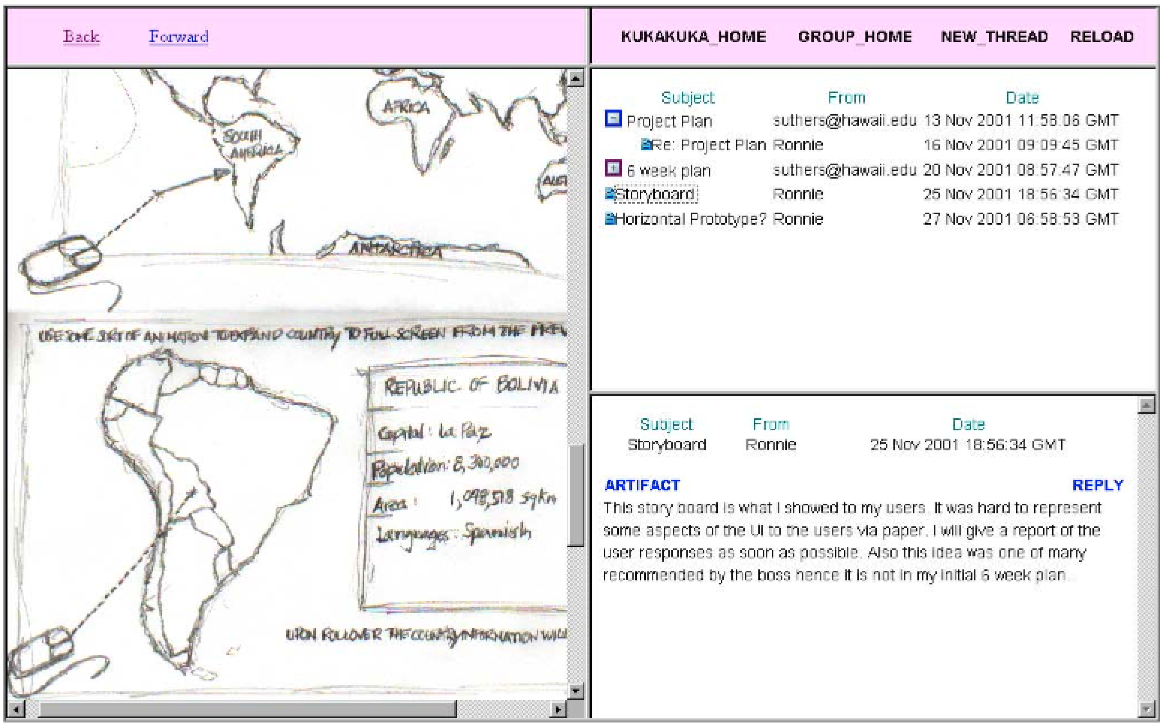
Ik weet niet of het bij jullie al besproken is, maar indoctrinatie kan volgens Frankena wel onder education vallen. Zolang het een positieve uitkomst heeft. Als voorbeeld hierbij de verkeersregels, die kun je er instampen zonder dat het kind weet waarom = indoctrineren. Toch valt het wel onder education want het is moreel wenselijk dat het de regels kent.

Themuch-discussed question of the relation of indoctrination to education is relevant here. Indoctrination appears to be one way in which the young might be made to acquire at least some of the dispositions Peters has in mind, though he may be meaning to rule out its use in education by his criterion that the learner must see, if only as a child, what he is doing and why it is desirable. What seems crucial in the debate about it, however, is not whether indoctrination passes this criterion but whether its use is desirable and morally unobjectionable[3]. Those who think it is never so tend to deny that indoctrination is a form of education, while those who think it sometimes is so tend to hold that indoctrination is a kind of education, even if they limit its use. This suggests that we rule indoctrination and other doubtful methods out of education by definition, if and only if we regarde them as undesirable or morally objectionable. Should we rule them out of education on any other grounds? To say

Figure 3-25: Annotation system developed by van der Pol (2006b). Relevant text blurbs on the right are highlighted and linked to the forum conversation on the left with symbols (e.g., '3' in the figure)

Purnell et al. (1991) found similar results in different settings. They studied the effects of splitting attention between technical illustrations and their descriptors on cognitive resources. Their results suggested that the format of technical illustrations was superior when descriptors were contained within the diagram, as cognitive resources were not required to integrate the descriptors and the diagram. This is referred to as the *split-attention effect* (Chandler & Sweller, 1992).

Mülhpfordt and Wessner (2005) developed ConcertChat, a chat communication tool in which participants can explicitly refer to other contributions or regions in the shared material. They found that explicit referencing leads to a more homogeneous discourse, i.e. to more homogeneous participation and more participation in parallel discussion threads. Stahl et al. (2006) reported similar results using ConcertChat in a math course, highlighting the importance of joint referencing for collaboration. The ConcertChat interface will be explained in more details in section 4.1.9, as it is one of the interfaces that will be used in the experiments presented in this thesis.



Subject	From	Date
⊟ Project Plan	suthers@hawaii.edu	13 Nov 2001 11:58:06 GMT
Re: Project Plan	Ronnie	16 Nov 2001 09:09:45 GMT
⊞ 6 week plan	suthers@hawaii.edu	20 Nov 2001 08:57:47 GMT
Storyboard	Ronnie	25 Nov 2001 18:56:34 GMT
Horizontal Prototype?	Ronnie	27 Nov 2001 06:58:53 GMT

Subject	From	Date
Storyboard	Ronnie	25 Nov 2001 18:56:34 GMT

This story board is what I showed to my users. It was hard to represent some aspects of the UI to the users via paper. I will give a report of the user responses as soon as possible. Also this idea was one of many recommended by the boss hence it is not in my initial 6 week plan.

Figure 3-26: Kükäkükä interface. Viewing a Thread's Contextual Artifact while Reading a Message (from Suthers & Xu, 2002). When a message in the thread's list is visualized then the image to which it is associated is automatically refreshed in the artifact's pane (right hand-side)

Suthers et al. (2002; 2003) examined how learners constructed graphical evidence maps, and how these maps were used by learners to support conversation through deixis in face-to-face and online conditions. They developed a system for artifact-centered discourse called Kükäkükä (see figure 3-26). The results showed that although external representations play important roles as resources for collaboration in both face-to-face and online learning, they are appropriated in different ways. In face-to-face collaboration, deixis was accomplished quite effectively through gesture. Suthers and colleagues explained how gesture is *spatially indexical*: it can select any information in the shared visual space, regardless of when that information was previously encountered or introduced. Online collaborators also used external representations for referential purposes, but through verbal deixis and direct manipulation rather than gestural deixis. Verbal deixis in the chat tool was *temporally indexical*: it most often selected recently manipulated items.

Bauer et al. (1999) also worked on the use of telepointers in remote collaboration. They used a repair task where a helper was guiding a worker to fix the problem. They showed that by using an augmented-reality telepointer a remote user can effectively guide and direct the helper's activities. The analysis of verbal communication behavior and pointing gestures indicated that experts overwhelmingly used pointing for guiding workers through physical tasks. While the use of pointing reached 99% of all cases, verbal instructions were used considerably less. In more than 20% of the cases, experts did not use verbal instructions at all, but relied on pointing alone instead. The majority of verbal instructions contained deictic references like 'here', 'over there', 'this', and 'that'. Because deictic references are mostly used in connection with and in support of gestures, this finding is a strong indication that participants naturally combined pointing gestures with verbal communication, in much the same way they do in face-to-face conversations.

3.6 Synthesis: deixis, gaze, conversation linearity

When people interact, deictic gestures help ground the conversation. Instead of using complex descriptions of elements of the context, conversants can simply point at things. This mechanism reduces misunderstandings, which are a natural product of human language. It also reduces the time required to reach a mutual understanding. Many studies report that, even at collaboration level, the possibility of using deictic gesturing or equivalent mechanisms, has positive implications. However, it is important to consider the following:

– Remote deixis was often implemented and experimentally evaluated using a telepointer. Although this solution allows efficient reference resolution, it has some problems. First, not all of the movements of such a pointer are associated with communicative intentions. Second, the observation of such 'gesturing' cannot guarantee the observer the interest, the attention, nor the intention of the emitter (e.g., movements might be involuntary or the mouse pointer can be left unattended). Third, this mechanism does not bind permanently specific utterances to pointer-indications on the shared plan. In other words, this association is made on-the-fly while moving around the pointer and therefore is not permanent. Finally, the resolution of the selection might be not enough for specific applications (e.g., the mouse pointer allows to select punctual zone, not rectangles or polygons).

- While the majority of experiments report positive effects for the use of a telepointer in remote collaboration, one study described null effects of the use of a telepointer in a controlled experiment (Fussell et al., 2004). In this study, the authors compared a face-to-face interaction to a remote collaboration where the helper could indicate parts of a model using a telepointer mechanism. Contrary to their hypothesis, they found that adding the pointer was not sufficient to improve performance over that of the video-only condition. They discussed possible explanations for this, such as the limited functionality of the cursor tool. They also hypothesized that the part of the task requiring effective pointing represented only a small percent of the total task and therefore was not influential.

- The experimental evaluation of the impact of Explicit Referencing on collaborative work at a distance has received little attention. We still know little of the interplay between the design of remote deixis mechanisms and other facets of human interaction. The way remote deixis is implemented may disturb the linearity of the conversation or either create a fracture between the referential workspace to which messages might refer and the conversation itself.

- Finally, few studies considered in detail how deixis is intertwined with gaze. When collaborators use a deictic gesture in co-located meetings they can also monitor with their gaze that their conversational partners perceived the gesture. However, depending on how a remote deixis mechanism is implemented then this visual acknowledgment might be difficult, or not even possible.

3.7 Gaze awareness

Scholars demonstrated how gaze is connected to attention and, in turn, to cognition. Grant and Spivey (2003) argued that attention is not an outcome of cognition but it can help restructure cognition. They report a study of participants solving a radiology problem. The subjects' eye movements were recorded over an image of a tumor. The authors showed that participants who successfully completed the task were more likely to look at the external part of the tumor image. Then in a second experiment, the authors changed the visual salience of this external part of the image of the cancer, thus affecting the completion outcomes. They concluded that eye movement patterns were related with problem solving processes. Comparable results were obtained by Pomplun et al. (1996).

Gaze is also used to marshal turn-taking. Argyle and Graham (1977) asked pairs of students to plan a European holiday varying the presence and the details of a map of Europe placed on the table between them. They found that the amount of time spent looking at each other dropped from 77 percent to 6.4 percent when the map was present. 82 percent of the time was spent looking at the map. Even when the map presented scarce details, subjects spent 70 percent of time looking at it. This finding suggested that participants were regulating their turns by looking at the shared object instead of looking at each other. Similarly, Vertegaal et al. showed that when someone is listening or speaking to individual, there is a high probability that the person looked at is the person listened or spoken to (Vertegaal et al., 2001).

Gaze is related to cognition or the management of interaction, and also more directly to collaboration. The perception of the gaze of the interlocutors is a great source of information for understanding what they are talking about or attentive to. Colston and Schiano (1995) studied how observers would rate the difficulty people had in solving problems using gaze information. Observers were basing their estimates on how long the participants observed would look at a particular problem and particularly how long her gaze would linger after being told to move on to the next problem. They found a linear relationship between gaze duration and the difficulty that was rated, indicating lingering as a significant factor. This suggests that collaborators use gaze information to infer the cognitive activities of a partner. Indeed this was verified by several research like the study of Brown-Schmidt et al. (2005), who examined how listeners circumscribe referential domains for referring expressions by monitoring the eye-movements of their partner as they engage in a referential communication task. They observed and confirmed linguistic theories according to which reference resolution is made through a series of heuristics. More interestingly for this thesis was the fact that the eye-movements of the emitter of a message are used by the listener to restrict the possible interpretations of a referent. The same finding was confirmed in a later study by Hanna and Tanenhaus (2003), and more recently by Hanna and Brennan (2007).

A more strict relation between gaze and collaborative work was demonstrated by Ishii and Kobayashi (1992). They showed that preserving the relative position of the participants and their gaze direction could be beneficial for cooperative problem solving. They used the ClearBoard system described above (see figure 3-11). To test the system they designed an experiment where they used a puzzle called "the river crossing problem", where missionaries or cannibals should reach the other side of a river according to a series of constraints. As solving the task is highly dependent on understanding where the opponent is looking, the use of ClearBoard had a positive impact on the task resolution. A similar setup was proposed by Monk and Gale (2002) (see figure 3-27). Their findings demonstrate a reduction of speech quantity and ambiguity. However, they did not find an improvement of performance over a control condition. We can hypothesize that the positive effects on language might be annihilated by a negative impact of displaying continuously irrelevant or intrusive information on the position of the partner's eyes. The same technique of

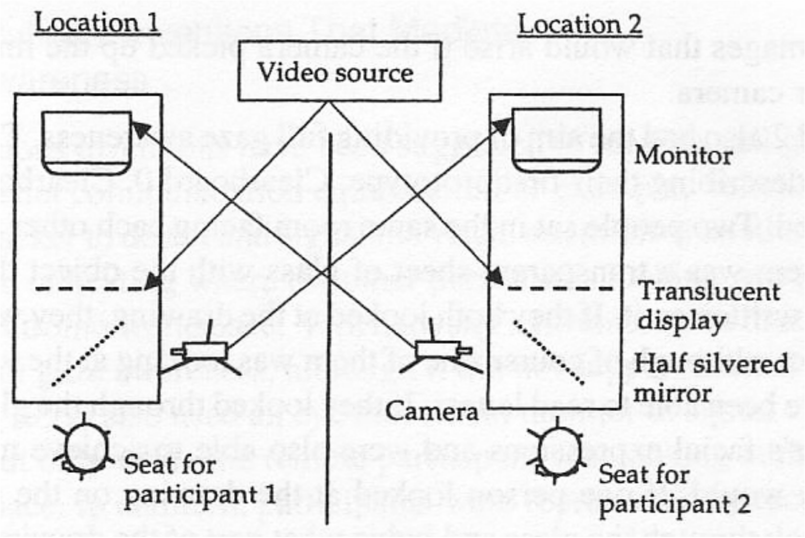


Figure 3-27: The GAZE awareness display designed by Monk and Gale for use in the experiment (2002). This system supported mutual and full gaze awareness

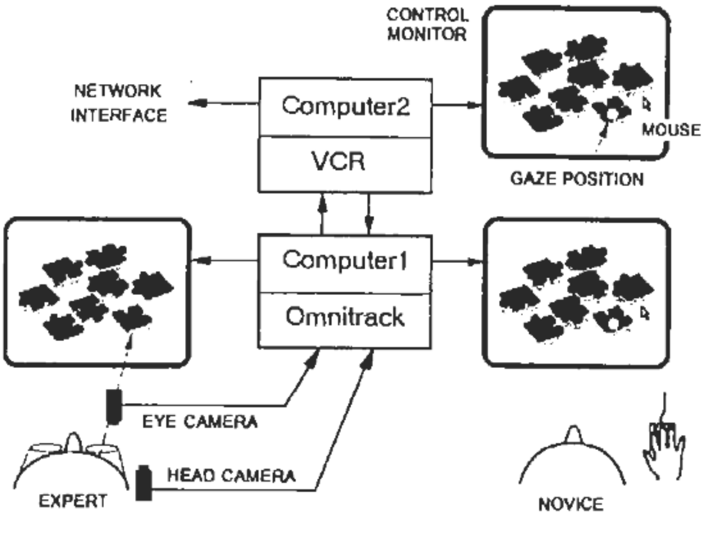


Figure 3-28: General arrangement of the experiment used by Velichkovsky (Velichkovsky, 1995)

employing a half-silvered mirror to optically align camera with video screen to enable eye contact was used by Buxton and Moran (1990), and named *video tunnelling*.

Finally, Velichkovsky (1995) highlighted the importance of transferring gaze information at distance for collaborative work. Two participants were asked to solve a puzzle collaboratively. One of them had access to the solution while the other was operating the moves on the target puzzle. While the participants shared the same visual workspace, one of them had access to the key but she could not rearrange the pieces. Velichkovsky manipulated the participants' communication features. In the control condition, the participants could only communicate via voice, while in a second condition, the gaze of the participant who had access to the solution was projected on the workspace of the other. In a final condition, the participant who had access to the solution could use a mouse pointer to show to the other the relevant parts (see figure 3-28). Both the experimental conditions, transfer of gaze position and pointing with the mouse, improved performance.

3.7.1 Attentive interfaces

The positive results obtained by this basic research on how people use gaze to work together stimulated designers to develop interfaces that could be controlled by gaze. Also, as gaze was observed to be strongly related to attention, there was a body of work concerned with the development of systems capable of taking into account and thus affecting the user's attention. Without any presumption of being exhaustive, I want to cite here some examples of gaze-based or attentive interfaces.

Zhai et al. (1999) developed a pointing method for computer input, dubbed MAGIC (Manual And Gaze Input Cascaded) pointing, that used eye-tracking (see figure 3-29). They presented an experimental setup where three different input mechanisms were compared: pure manual, pure gaze, and a mixed approach. The authors' first claim was that a pure gaze interaction mechanism is unnatural as it overloads a perceptual channel. The authors tested the different input mechanisms with 36 subjects. Subjects using the gaze only pointing performed worse than those using the pure manual pointing mechanism. The best performance was achieved with the mixed approach.

Maglio et al. (2000) developed SUITOR, Simple User Interests Tracker, an attentive system that payed attention to what the user is looking at and through probabilistic models inferred what her interests might be. Then this information was used to create a peripheral awareness around the topic of interest.

Vertegaal developed the GAZE groupware system to overcome the problems of mediating multiparty communication, showing examples of how multiparty communication using video conferencing is not necessarily easier to manage than using telephony (Vertegaal, 1999). Single-

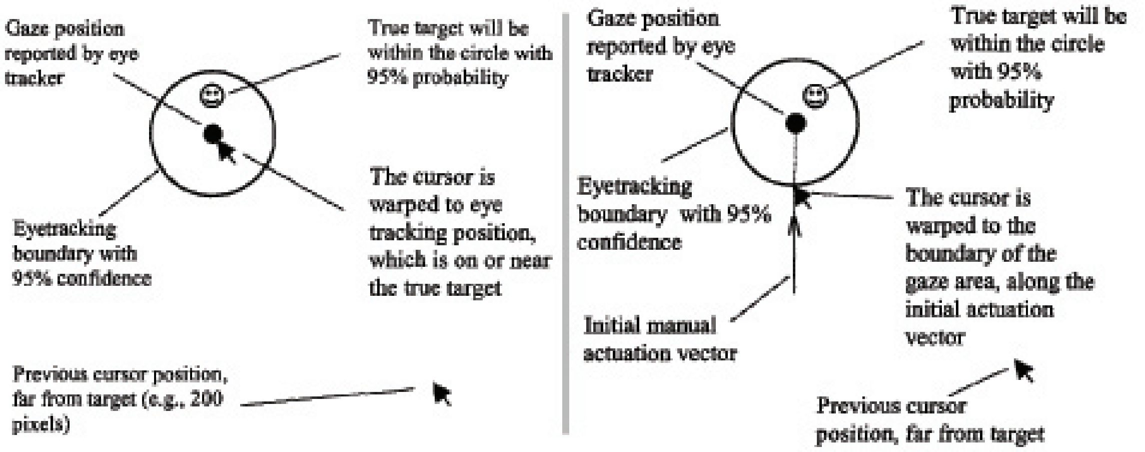
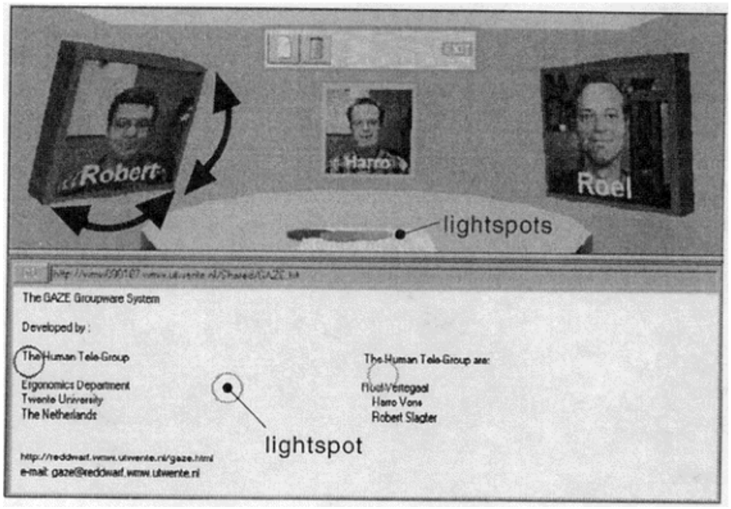

Gaze position reported by eye tracker
True target will be within the circle with 95% probability
The cursor is warped to eye tracking position, which is on or near the true target
Eyetracking boundary with 95% confidence
Previous cursor position, far from target (e.g., 200 pixels)
Gaze position reported by eye tracker
True target will be within the circle with 95% probability
Eyetracking boundary with 95% confidence
The cursor is warped to the boundary of the gaze area, along the initial actuation vector
Initial manual actuation vector
Previous cursor position, far from target

Figure 3-29: The liberal MAGIC pointing technique on the left (Zhai et al., 1999): cursor is placed in the vicinity of a target that the user fixates on. On the right, the conservative approach: an intelligent offset is used



The GAZE Groupware System

Developed by :

The Human Tele-Group

Ergonomics Department
Twente University
The Netherlands

The Human Tele-Group are:

Roel Vertegaal
Harro Vons
Robert Slagter

http://reddwarf.wmw.utwente.nl/gaze.html
e-mail: gaze@reddwarf.wmw.utwente.nl

Figure 3-30: The GAZE groupware system designed by Vertegaal (1999). Personas rotate according to where users look

camera video systems do not convey deictic visual references to objects or persons. To overcome most of the limitations of previous approaches, he implemented a virtual meeting room where the camera feed of each participant is represented in a moving panel that replicates the movements of the head of the real participant. This plus a lightspot on the shared workspace gives a sense of gazing at other participants and gazing at the objects used during the interaction (see figure 3-30). Although the system was described in great detail, no evaluation was reported. Vertegaal argued that a problem in designing *mediated systems* is conveying the least redundant cues first. The central issue of supporting collaborative work at distance is that regardless of whether audio or video is used one should provide simple and effective means of capturing and metaphorically representing the attention participants have for one another and their work.

In a similar approach, Oh et al. (2002) developed look-to-talk, a gaze-aware interface for directing a spoken utterance to a software agent in a multi-user collaborative environment. Using a controlled experiment, they showed that their prototype was a natural alternative to speech and could estimate the focus of attention of the participants.

3.7.2 Summary: opportunities and issues of gaze-based interfaces

It is a shared conviction that eye-based interfaces offer enormous potential for efficient human-computer interaction, but also that the challenges for a proficient use of this technology lie in the difficulty of interpreting eye movements accurately and displaying them in a smart way. Just as it is difficult to infer comprehension from users' speech, eye-movement data analysis requires a fine mapping between the observed eye movements and the intentions of the user that produced them (Salvucci, 1999).

Additionally, Wood et al. (2006) express caution regarding eight issues that are fundamental when designing attentive systems. I report four of them here as they summarize the discussion reported in this section well.

– First of all, there is no consensus on *what is* attention. Many models are proposed that tend to represent attention as a spotlight or as a 'coherence field'. Most research tends to assimilate attention with a visual search through eye-gaze. However this is not necessarily correct because a person can look at some point of an interface while thinking to something else.

– Second, attention is difficult to measure. Vision has a selective nature. The implication is that direction of gaze is not necessarily a synonym with focus of attention and therefore studies will need to validate what the focus of attention is through further research.

– Third, it is imperative to understand how graphical displays interact with attention. Larkin and Simon (1987) demonstrated that the way an external representation encodes information is critical to how easy it will be for the user to find relevant information: pictures and diagrams have advantages over textual descriptions. This assumption was questioned by Underwood et al. (2004), who argued that the ease of recognition of relationships from a picture was not reflected in fixation durations. Therefore, Underwood and colleagues concluded that richer representations of information in pictures require extensive encoding durations which are comparable to the encoding of information from text.

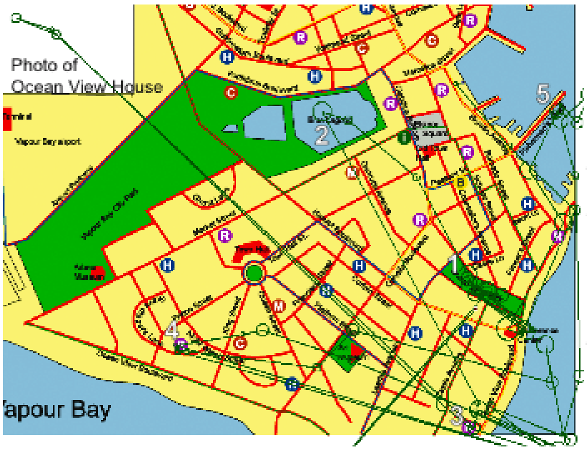
– A fourth point of interest concerns the potential effects of introducing artificial feedback in systems designed to monitor the user's attention. It has been demonstrated that the introduction of artificial feedback loop can, in some circumstances, cause variables that are generally highly correlated to become decoupled.

In conclusion, the knowledge of where collaborators are looking has a tremendous value in sustaining collaborative work at a distance. However, as we have seen for deictic gestures, this information alone is difficult to associate with cognitive processes or simply to use as an interaction mechanism for driving interfaces (e.g., the *Midas touch* problem). The combination of gaze with other users' interaction information offers great potential.

3.8 Intelligent interfaces: mixing language models and multimodal user's input

During the last decade, many applications have been proposed that combine language models with other sources of information such as gaze or user's interaction (e.g. clicking on specific parts of the interface). The basic idea of these approaches was that one channel of information could help decode or complement the information provided by the other channel. Some of these projects are relevant to this thesis.

Qvarfordt et al. (2005), developed a system called RealTourist that allowed a tourist to consult a remote tourist consultant with shared visual information on a computer screen. The system overlaid the tourist's gaze locus onto the consultant's view of the shared workspace (see figure 3-31). The authors conducted qualitative observations of the system in use and derived some interesting conclusions: eye-gaze carries deictic and spatial reference information, the display of which might reduce effort of frequent referencing. Additionally, the authors noticed how eye-gaze reflects a listener's interest and can be used to judge whether or not to continue with the current conversation topic. They also found that providing this information reduced the ambiguity and increased redundancy in communication. However, these conclusions are purely qualitative and need to be verified under a proper quantitative paradigm.



a1 T14: well I'd like a nice place where I have a nice view

a2 TC A: okay so we have for example the Lobert House [Ocean View House 3] it's right along the ocean it has seafood and ... it's a very novel fish [restaurant] ... they have a varity of fish and ...

a3 T14: okay what about something near the Fisherman's Wharf

Figure 3-31: Example of conversation from RealTourist (Qvarfordt et al., 2005). Green lines represents scanpaths of the tourist consultant, and indicate changes of focus and different interest levels. The vertical line in transcription indicates the time periods used for the gaze fixation trace

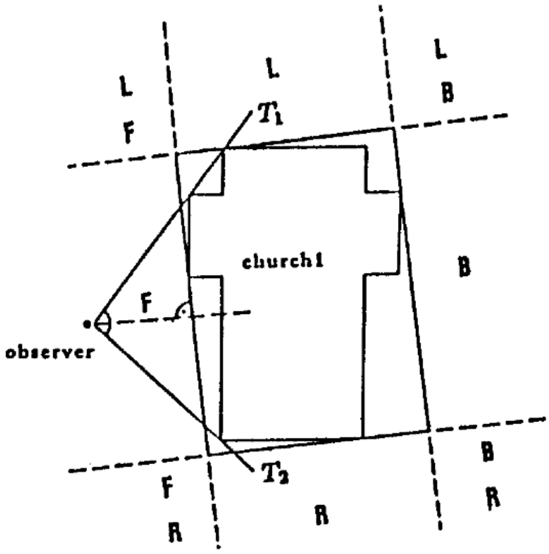


Figure 3-32: CITYTOUR answered natural language queries based on the user's position in the virtual tour of the city (André et al., 1987). The computational model used a delineative rectangle based on the observer's position

André et al.(1987) developed a computational semantics of natural language expressions describing spatial relations between physical objects. They tested their system using CITYTOUR, a German question-answering system that simulates aspects of a fictitious sightseeing tour through an interesting part of a particular city. The user of CITYTOUR could send natural language queries to the system that were evaluated and answered depending on the her placeholder position in the city (see figure 3-32). Although the model was described in great detail, it was not evaluated in a proper user study. A similar project was conducted by Lokuge et al. (1995). They were interested in helping users of a Geographical Information System to formulate queries. They developed an interface called GeoSpace that allowed the user to formulate vocal queries about spatial features in the Boston area. The interface was conceived as a series of linked layers containing the relevant information. Each layer's transparency was regulated by the user's interaction over the map. If the user zoomed to or expressed an interest in a particular area the layout and the information visualized was adapted accordingly.

Campana et al. (2001) also worked on an computational model for reference resolution employing eye-movements. They took inspiration from a human-to-human interaction, where participants under-specify their referents, relying on their discourse partners for feedback if more information is needed to uniquely identify a particular referent. By monitoring the eye-movements of the user, they aimed at improving the performance of a spoken dialogue system by observing referring expressions that were under-specified according to a linguistic model. The paper did not present a systematic evaluation of the system.

The work of Ou and colleagues on Shared Visual Space suggested an interesting application: the idea of using the eye-movements of the participants over the different parts of the interface, and an automatic parsing of the exchanged message to drive the switch of an multi-view collaborative system (Ou, Oh, Yang, & Fussell, 2005; Ou, Oh, Fussell, et al., 2005). Basically, the system did not need user's input to change the camera view but this mechanism was provided by an automatic parsing of the participants' language. The study showed that the gaze movements were systematic during the task and could be predicted on the basis of what the participants were saying. Their point was that given the limited resources available to sustain work at a distance (e.g., bandwidth is expensive), it is an advantage to be able to switch view dynamically. However, as noted by Kirk (2006, p. 30):

> ...this largely ignores the complexity of actually parsing spoken language, and brushes over the large amount of inaccuracy that the presented system demonstrated. The technology design also fundamentally assumes that visual saccades and general visual attention follows changes in speech pattern and not the other way around, which unless empirically tested and demonstrably shown to not be an issue of concern is potentially going to significantly hamper use of such a technology.

3.9 Summary and Conclusion

The objective of this chapter is to describe the various threads of research which are important for framing the research questions and the experiments that constitute the core of this thesis. I detailed many of the issues that brought researchers to design various technologies to support remote collaboration. I structured this discussion using the framework presented in the introduction, and for each approach I presented relevant critiques that emerged from the literature. It is now worthwhile to conclude this chapter by briefly summarizing the major issues discussed.

- *Efficient communication uses a plethora of channels.* Linguistic research suggested that communication relies on more than the expression of verbal actions. Conversational partners take advantage of the availability of different modalities of communication, mixing verbal and non-verbal actions to achieve maximum clarity with minimum effort. These non-verbal behaviours can be essential in interpreting verbal information and can provide an additional support for interaction.

- *The less the better.* Researchers initially strived to support remote collaboration by designing technological solutions aimed at imitating face-to-face interactions. This initial quest explored strategies to facilitate communication between the remote sites offering video-links. Ethnographical research in this area revealed that the more we attempt to enhance the links between the remote environments, the more we may exacerbate difficulties for the coordination of action. In this perspective, I argue, together with other researchers, that we should avoid the *imitation bias*. Interfaces at support of remote collaboration should implement mechanisms which are modeled upon face-to-face processes. These should be functionally equivalent to embodied communication strategies, and they should strive for the same efficiency. However, as their context of deployment is completely different from co-located interactions, they should translate in a different manner their underlaying communication principles.

- *Communication is a cooperative process.* Human communication is adaptive in its production and interpretation. This implies that a maximally shared visual environment is not required to obtain the best comprehension efficiency. One of the challenges of sustaining collaborative work at a distance is to identify the minimum communication channels necessary to provide remote collaborators with an optimal shared visual context (and not a maximal shared visual context).

- *Video links are not enough.* Research findings have been inconsistent about the benefits of providing video links between the remote sites for sustaining collaborative work. The usefulness of video-information is related to the task at hand. There is evidence that providing video is useful in particular object-focused interaction, where collaborators need to coordinate their efforts around a common object of interest. Systematic observations of teams using video-links in shared workspaces also demonstrate how video technology may introduce distortion of perspectives and other biases that could hamper collaboration.

- *Deixis is a basic pragmatic need of collaboration.* When collaborators interact over a shared workspace, they can greatly benefit from the ability to gesture. In particular, deictic gestures are one form of gesturing that is performed very often during design meetings and that is proven to be useful for disambiguating references, making communication more efficient. Deictic gestures are in their essence multi-modal. They join elements in the visual field to specific linguistic contributions and are perceived and acknowledged through gaze.

- *A Telepointer is not remote deixis.* Studies often implemented remote deixis through the means of a telepointer. However, this solution is biased because it detaches the gestural activity from the corresponding comunicative intentions. Therefore, this mechanism does not guarantee intentionality, proper resolution or permanence of the gestures. More specific solutions to support remote deixis have been introduced in chapter 2, but these have not been the object of quantitative evaluations. This is one of the objective of this thesis.

- *Explicit Referencing design can impact conversation linearity.* Different design implementation of Explicit Referencing can organize the messages according to a spatial criteria, which might affect the conversation linearity, or following a time criteria, which might create a split-attention between the shared workspace and the conversation. This dimension has not been explored in the literature. Therefore, to fully establish the benefits of explicit referencing on collaborative work at a distance, a more systematic evaluation of this mechanism would be needed. This, in turn, would shed light on the best ways in which such systems can be constructed and deployed.

- *Deixis and Gaze.* Finally, linguistic theories suggests that remote deixis is intertwined with gaze awareness when face-to-face gestures can be perceived and acknowledged by recipients using gaze. This is not possible at distance, at least not with commonly available technology. A second objective of this thesis is to better understand how these two mechanisms are related and to develop strategies to better exploit them in supporting collaborative work at a distance.

Chapter 4

Location-Based Annotations

The objective of this chapter is to analyze Collaborative Annotation Systems that implement in different ways the mechanism of Explicit Referencing. I describe these systems highlighting common features and proposing a classification based on nine dimensions, without any presumption of being exhaustive of the available systems and their differences.

4.1 Collaborative Annotation Systems: a classification

In these recent years, numerous applications have emerged offering in various degrees and forms the possibility of connecting information to specific point in the actual space or in the space of a shared display. I refer to them as Collaborative Annotation Systems, or CAS in short. These systems implement the interface mechanism of Explicit Referencing that I discussed in the previous chapter. There is no established classification of CAS tools. However, I identified nine characteristics that can help to categorize them: (a) the group size proposed by their scenario of use; (b) the temporal span over which interaction should occur; (c) the geographical area covered by the mapping support; (d) the degree of immersion in the actual space for which the service is designed; (e) the way location is acquired; (f) the way spatial information is handled by the application, (g) the way the content is anchored to the shared workspace; (h) the criterion under which the messages are organized by the application; and finally, (h) the way the comunicative content is anchored to the shared map. I will detail these characteristics below.

(a) *The group size* defines how many users interact with the service: on one end we find **individuals or small groups**; on the other extreme, we find larger groups or **communities**. In the former case the participants probably know each other and have a common goal to achieve using the system, while this is not necessarily so in the latter case. Community members have a multiplicity of intentions and the identity of each member might not be known by all the other members. In the former case, the tool is designed to support a specific scenario for individuals or small groups. In the latter case, tools strive for maximum adaptability to the different members' goals.

(b) *The time-span of the interaction*. CAS tools can either support **asynchronous** interaction (e.g., the exchange of an SMS) and **synchronous** interaction (e.g., much like a chat interchange). The term synchronous interaction will be loosely applied to situations in which participants exchange messages in real-time.

(c) *The geographical scale* for which the service is designed. Some CAS are developed to serve the communication needs of a community which lives in a particular region. In this case the geographical support will provide limited map coverage for the region of interest, namely a **building** or a specific **neighborhood**. Alternatively, when the CAS provides annotation functionalities to a wider audience then map coverage will increase to the **city** level or to the **country** where the service is provided (and eventually to the world).

(d) *The degree of immersion* in the actual space describes whether the service/platform has been developed for **ubiquitous** or **fixed use**. When the geographical position of the user does not play a role in the production or retrieval of the information, I talk about a 'fixed use' system, using which the user can contemplate space. When the actual position matters, then I talk about an 'ubiquitous' system, using which the user can experience the space.

(e) *The way location information is acquired*. Some CAS systems are designed for mobile devices, which can collect location information **automatically** from network information. When this is not possible, users are meant to input location information **manually**. Automatic location gathering concerns both the message input phase and the retrieval phase of the available content. Various combinations of these two solutions are also possible.

(f) *The way location is described*. As discussed in chapter 2, this corresponds to the place/space distinction. Location can either be **discrete** such as places names, or **continuous** using a coordinate system. Of course, in the latter case the coordinates will refer to the virtual or physical environment, or eventually to the relative coordinates in the map hosting the interaction.

(g) *The way the content is anchored to the shared workspace*. Some CAS allow users to relate annotations to a **point** of the shared workspace while other to an area such as a **polygon** (a simple rectangle or a more complex shape).

(h) *The organization criterion*: this is the strategy used to organize the information in the collaborative system both at storage level and presentation level. Messages can be presented and retrieved using a **time** criterion or they can be grouped by **content** or context to which they refer. Chat messages are typically displayed using a time-criterion, while in a forum they can also be organized according to the thread to which they refer.

(i) *The presentation of the referencing link.* This connection can either be presented visually or through other multi-modal interaction mechanisms, such as reminders, or audio alerts. Different mechanisms can be used to visually present the connection between the communication and the location to which the message refers: notes can be **overlaid** over the shared workspace or they can be displayed **beside** of the workspace, using links or symbols to match anchor and description.

These categories are not orthogonal dimensions but enable me to distinguish the systems that will be presented in the rest of this text. Figure 4-1 resumes visually the dimensions. The list of applications below should not considered exhaustive. It only provides examples of applications that illustrate the different types of interface defined by the criteria above. Table 4.1, at page 125, presents the classification in respect of the categories above.

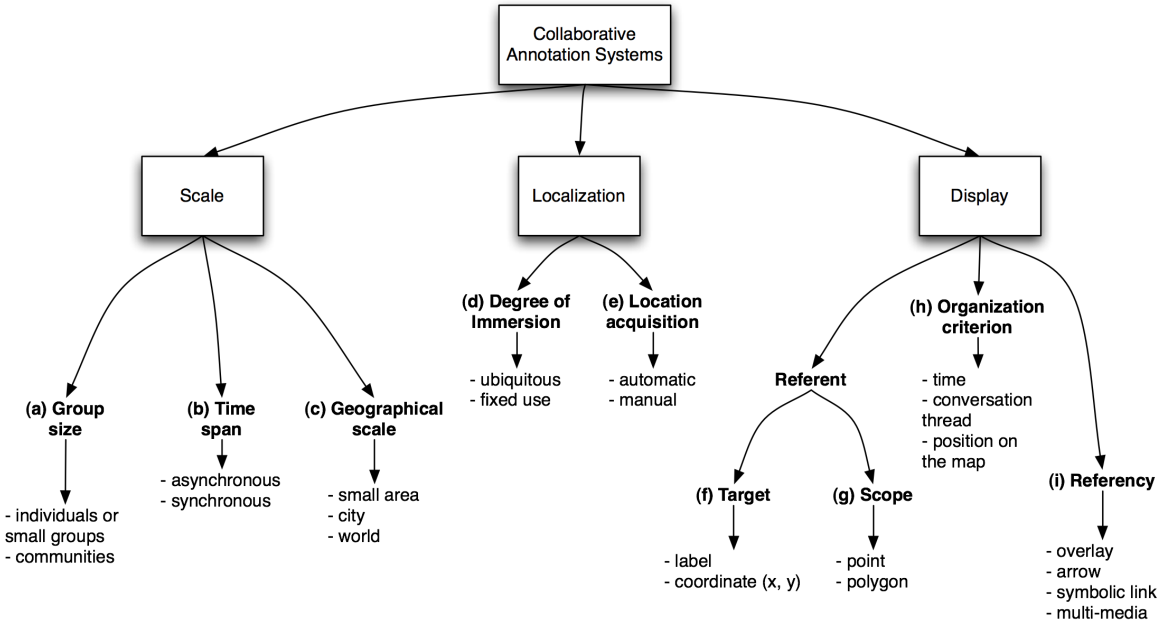
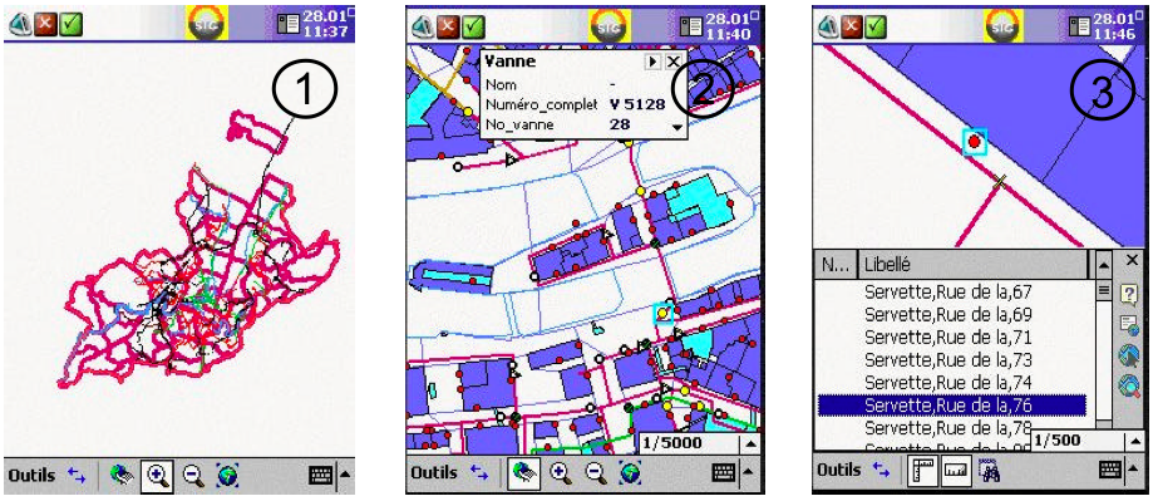


Figure 4-1: Graphical summary of the categories that will be used to describe the differences of the CAS

4.1.1 (a) The group size

CAS systems can be designed for specific applications, in order to sustain communications between few individuals. Else, systems can target larger-scale populations. The applications can provide a specific protocol for interaction, like a customized vocabulary to be used for specific tasks in the former case, while applications in the latter group will target generic applications, often supporting multiple tasks and striving for integration with other existing services or platforms.



Vanne
Nom -
Numéro_complet V 5128
No_vanne 28

N...	Libellé
	Servette,Rue de la,67
	Servette,Rue de la,69
	Servette,Rue de la,71
	Servette,Rue de la,73
	Servette,Rue de la,74
	Servette,Rue de la,76
	Servette,Rue de la,78

Figure 4-2: PDA interface of GeoConcept. (1) The overview of the map of Geneva; (2) specific information on the electrical network are overlaid on the cadastral map and a small annotation that was left by one of the technicians; (3) location-specific queries can be run against a remote gazetteer

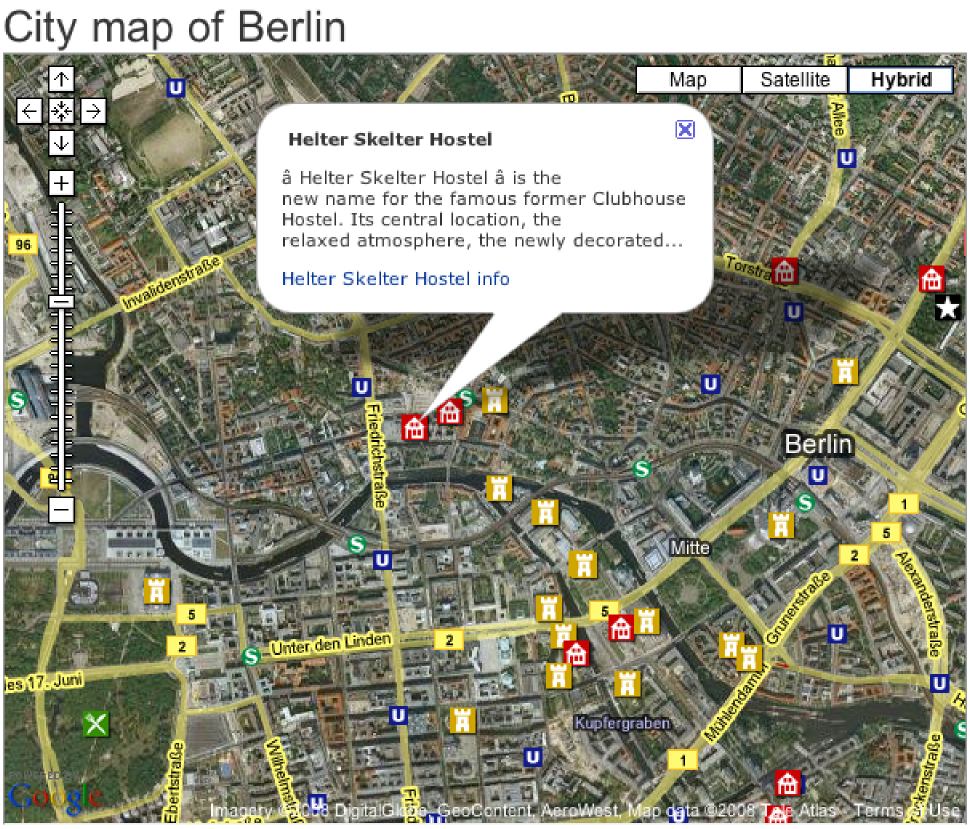
An example of CAS application in the "small group" pole is GeoConcept, a tool for PDA developed by the eponymous French company[1], provides a geographical information system to industry professionals that maintain geographically distributed resources. In Geneva, employees of the public electric company[2] use GeoConcept to annotate their maintenance activities and coordinate with the dispatch unit. The tool shows their position on the screen of the PDA, in addition to specific information about the electrical network at the location. The repair specialist can retrieve and update this information to coordinate with the control room and with other colleagues while on the field. The application allows the technician to attach text annotations, as small sticky notes, to the cadastral maps for other colleagues who will continue their work at later times. Figure 4-2 shows screenshots of the PDA interface of GeoConcept. The location of every object in this application is stored with geographical coordinates.

Conversely, world66[3] is a CAS application that is not designed for supporting the communication within a specific group. Instead, it gathers contributions from tourists from all around the world. In fact, every user can use this system to report locations that she found interesting, or

[1] GeoConcept SA, Bagneux, France. See `http://www.geoconcept.com/` , last retrieved January 2008.

[2] See `http://www.sig.ch` , last retrieved January 2008.

[3] See `http://www.world66.com/`, last retrieved January 2008.



City map of Berlin

Helter Skelter Hostel

â Helter Skelter Hostel â is the new name for the famous former Clubhouse Hostel. Its central location, the relaxed atmosphere, the newly decorated...

Helter Skelter Hostel info

Figure 4-3: Map tool of world66. Icons indicate available annotations organizing them into several categories

that were to be avoided, during a recent trip. She can upload a picture of the place and a brief description of the resources available there. More importantly, the user needs to give specific coordinates to link the information given to their geographical information system. Usually this is done by pointing to the coordinates of interest on the map. The collected information is potentially used by a traveller to look for specific resources before going to the place. Through the web site interface, the user can see a map of the place to be visited and icons illustrating available annotations in specific categories (see figure 4-3). Additionally, the results of the queries of the system can be exported in various formats and uploaded on mobile devices for use in the field.

4.1.2 (b) Interaction time-span

While some CAS systems are designed to support synchronous interactions like people meeting briefly online for solving a problem in real-time, others are meant to support interactions spanning over days or weeks or without any precise time-constraint at the time the conversation is initiated. Below, I present two representative applications that are used for synchronous collaboration at distance and using desktop computers.

A tool that supports synchronous conversations over a shared map is called MapChat, an extension of a geographical information system to sustain map-based interaction that was developed by Hall and Leahy (2008). This tool allows one to browse complex vector maps while maintaining a chat conversation on a side. When a specific message relates to a particular region of the map, the user can make this link explicit by using a specific command. Then the message appears overlaid to the map as a "bubble" (see figure 4-4).

A counter-example in this category is given by WikiMapia[4], a service similar to dismoiou, discussed in section 4.1.4. The goal of this service is to annotate places for touristic or documentation purposes. Particularly, it combines a map service such as GoogleMaps and a wiki documentation service like Wikipedia, therefore providing an encyclopedic description for relevant places on Earth. As entries are not meant for a specific interaction period, the service was designed for asynchronous interactions.

4.1.3 (c) Geographical scale

Some CAS are designed to offer annotation support to a small geographical area, which may be a campus, a particular neighborhood or even a specific building. Other CAS are designed for annotating a much wider area such as a city, a country or the whole world. Most of the examples of CAS for fixed use presented in this section fall in this latter category (e.g., the WikiMapia service above). Fewer CAS are developed for specific use on the field.

An example of CAS for small area annotation is FieldMap (see figure 4-5), an application that was specifically designed for archaeologists' work on the field (Ryan, 2005). The software runs on PDAs and allows the user to upload the map of the excavation for consultation in the field and for note taking. The location acquisition is automatically done through a GPS device paired with the PDA[5].

4.1.4 (d) Degree of immersion

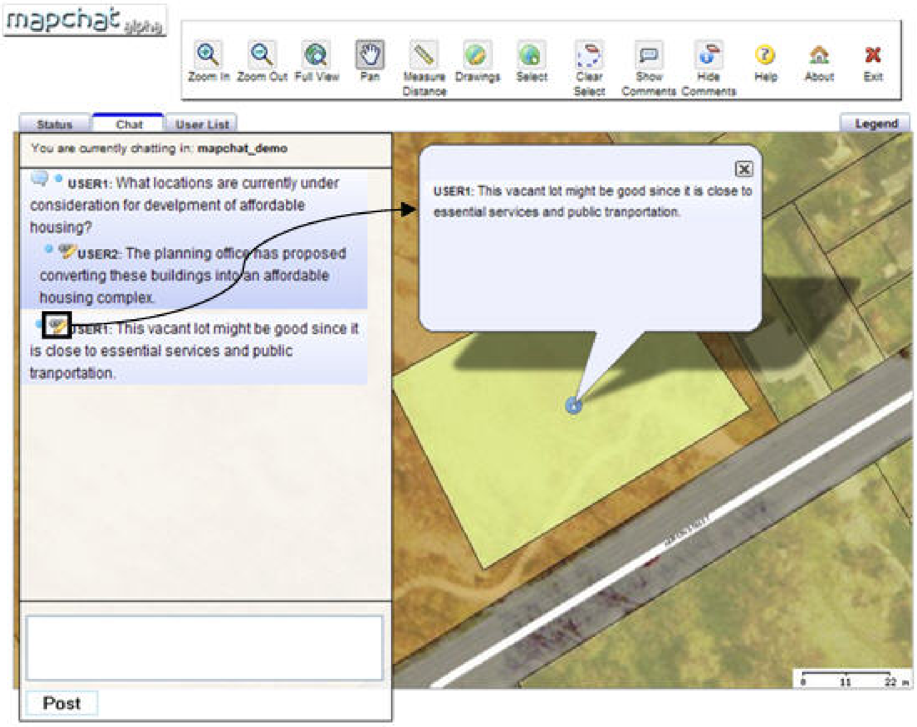
The key principle of mobile technology is that *location matters*. The prototypical example of such devices is a mobile phone, that allows people to communicate while being non co-located. As a side effect, mobile devices and their interface strive for minimalism. They need to be light and to fit in a pocket in order to be considered portable. On the other end of this spectrum, we find applications for which the actual location of the user does not play a role in the service being offered. These are also feature-rich applications that aim at offering a compelling user experience.

An example of a mobile CAS application is iMapFan[6]. This tool was developed for the Japanese

[4] See `http://wikimapia.org/`, last retrieved January 2008.

[5] See `http://www.mobicomp.org/FieldMap/`, last retrieved April, 2008.

[6] See `http://www.mapfan.com/imapfan/`, last retrieved January 2008.



You are currently chatting in: mapchat_demo

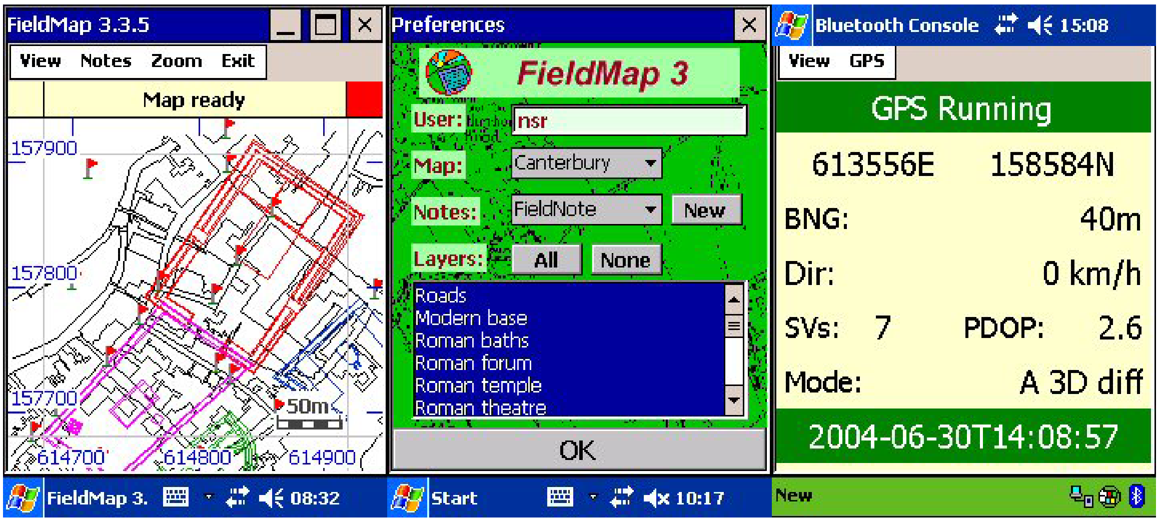
USER1: What locations are currently under consideration for develpment of affordable housing?

USER2: The planning office has proposed converting these buildings into an affordable housing complex.

USER1: This vacant lot might be good since it is close to essential services and public tranportation.

USER1: This vacant lot might be good since it is close to essential services and public tranportation.

Figure 4-4: A screenshot from MapChat: submitting chat messages linked to selected objects in a map (from G. B. Hall & Leahy, 2008)



Map ready

User: nsr

Map: Canterbury

Notes: FieldNote New

Layers:

Roads
Modern base
Roman baths
Roman forum
Roman temple
Roman theatre

GPS Running

613556E 158584N

BNG: 40m

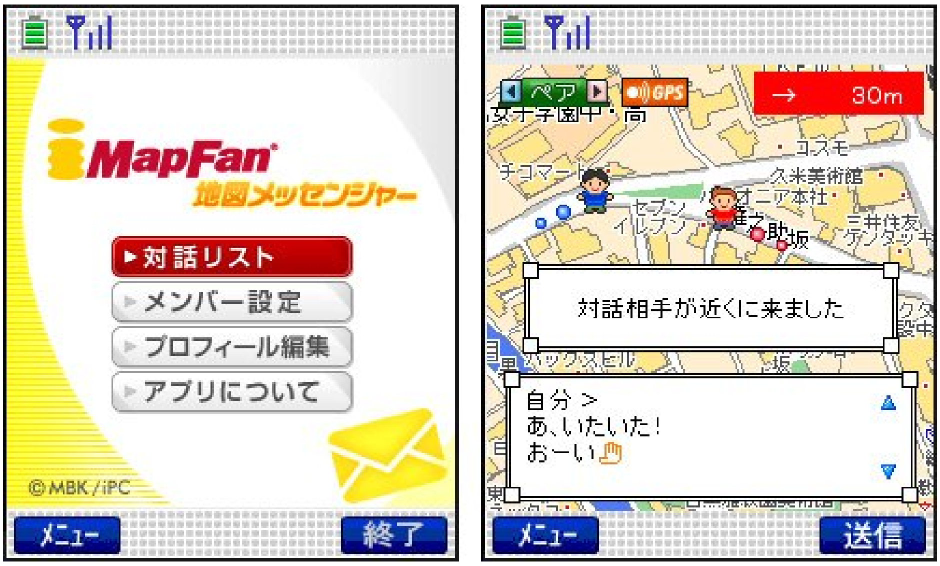
Dir: 0 km/h

SVs: 7 PDOP: 2.6

Mode: A 3D diff

2004-06-30T14:08:57

Figure 4-5: Screenshots from the interface of FieldMap (from Ryan, 2005)



対話相手が近くに来ました

自分 >
あ、いたいた！
おーい

Figure 4-6: Screenshots of the iMapFan interface. On the left hand-side, the main menus, while on the right hand side, two users chatting. The landmark is displayed with an avatar icon, while the last messages are displayed in the white boxes below

mobile market and emerged in the ecology of *i-mode* services[7]. In Japan, SMS messages are exchanged at high frequency by teenagers enabling a form of *mobile chatting* with a group of peers (see section 2.3.4). It allows the user to exchange short text messages synchronously with one or few friends. The iMapFan is unique in that each messages is enriched with the location of the emitter. The exchanged messages are displayed on the recipient(s)'s interface as a sticky note pinpointed to a map at the emitter's exact coordinates. Figure 4-6 shows a screenshot of the application.

A counter-example concerning this mobility aspect can be given by the community portals that allow users to express location-based recommendations. DisMoiOu[8] is basically a community-edited wiki[9]. Each article of this wiki describes a geographical place. The main interface presents a map of the city with navigation mechanisms. The annotations appear as pinpoint icons on this map (see figure 4-7). Clicking one of the pinpoints opens a pop-up window showing the annotation and its social rating. This system can be used to indicate the best pizzeria in town, for instance. Messages are retrieved using the map as a primary filter mechanism. In other words, users looking for recommendations in a certain city will first filter the dataset restricting it to a geographical region, then they will use a keyword search to select a specific category among the articles present in that region. The interaction scenario is designed for a user planning a future

[7] i-mode is a wireless Internet service popular in Japan and is increasing in popularity in other parts of the world. See `http://en.wikipedia.org/wiki/I-mode`.

[8] See `http://dismoiou.fr/`, last retrieved January 2008.

[9] A wiki is software that allows registered users or anyone to collaboratively create, edit, link, and organize the content of a website.

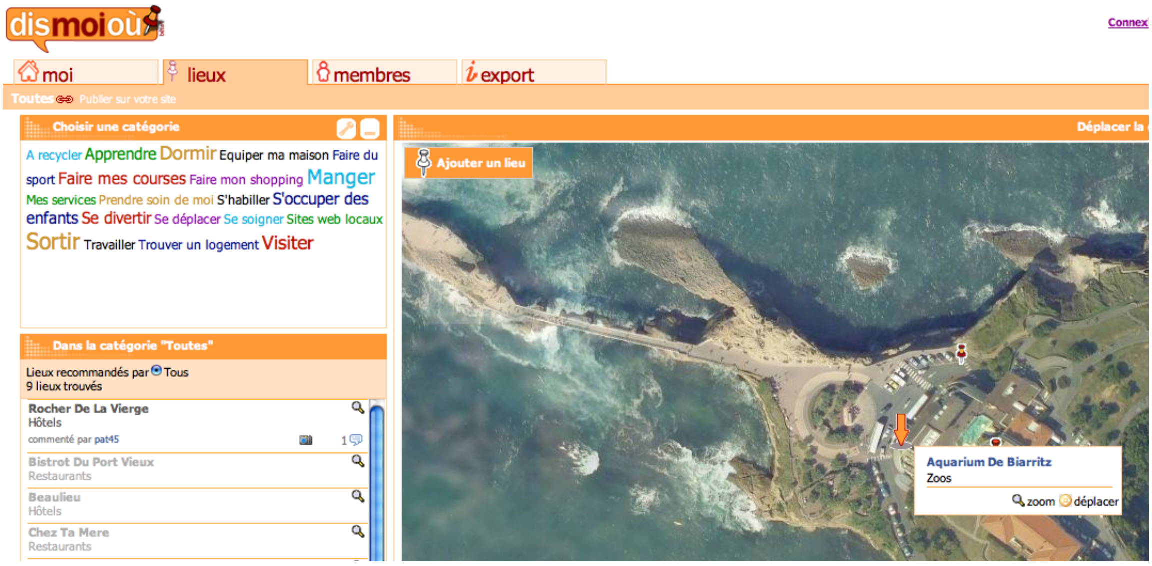

dismoioù
Connex
moi
lieux
membres
export
Toutes
Publier sur votre site
Choisir une catégorie
A recycler Apprendre Dormir Equiper ma maison Faire du sport Faire mes courses Faire mon shopping Manger Mes services Prendre soin de moi S'habiller S'occuper des enfants Se divertir Se déplacer Se soigner Sites web locaux Sortir Travailler Trouver un logement Visiter
Dans la catégorie "Toutes"
Lieux recommandés par Tous
9 lieux trouvés
Rocher De La Vierge
Hôtels
commenté par pat45
1
Bistrot Du Port Vieux
Restaurants
Beaulieu
Hôtels
Chez Ta Mere
Restaurants
Déplacer la
Ajouter un lieu
Aquarium De Biarritz
Zoos
zoom
déplacer

Figure 4-7: Dismoiou web-portal. Annotations are visualized as pushpins over the map. Clicking on a pin displays the content of the annotation

trip from a fixed location such as the home or the office.

4.1.5 (e) Location acquisiton

Many ubiquitous CAS capture the position of the user using network information or specific positioning technologies like GPS devices. More often, applications let the user select manually the zone of interest over the map. For instance, iMapFan, described above, relies on the GPS device internal to the phone.

4.1.6 (f) Referent description

The location to which an annotation refers to can be either described with labels such as the place name or with continuous measures such as the geographical coordinates. For instance when geographically-linked annotations are generated with a mobile device, each extra keystroke necessary to interact with the service is a burden. Developers have tried to simplify the interaction mechanism of these applications by automating the steps necessary to associate the message to its geographical context. In fact, some CAS applications use automatically available positioning information such as Cell-ID[10], to establish this mapping. For instance, ZoneTag is an application developed at Yahoo! Research[11] that allows the user to publish a picture taken with a mobile phone on a picture sharing community portal. The picture is usually enriched with a descriptive

[10] This is the unique identifier with which each cellular tower is identified worldwide. Each mobile device when turned on is in contact with one of such antennas. This information is used for a rough positioning of the device. See `http://www.cellspotting.com/`, last retrieved January 2008.

[11] See `http://zonetag.research.yahoo.com/`, last retrieved January 2008.

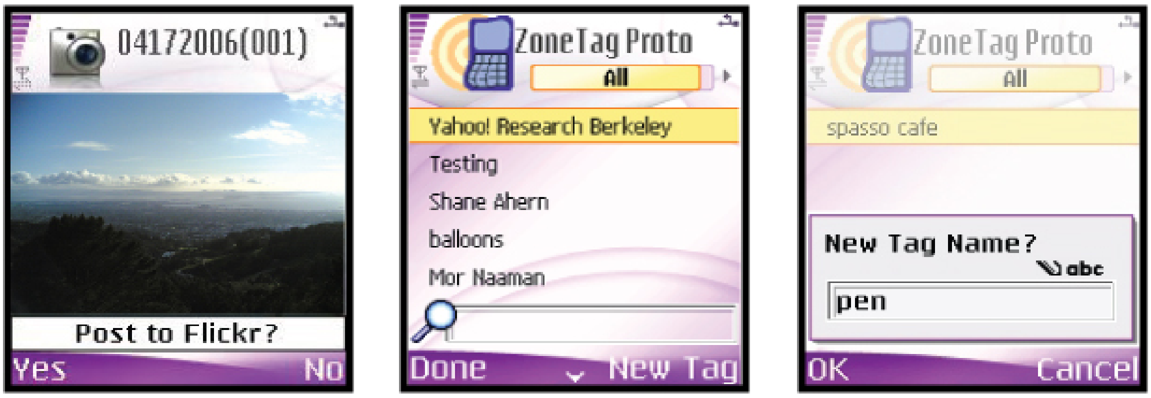
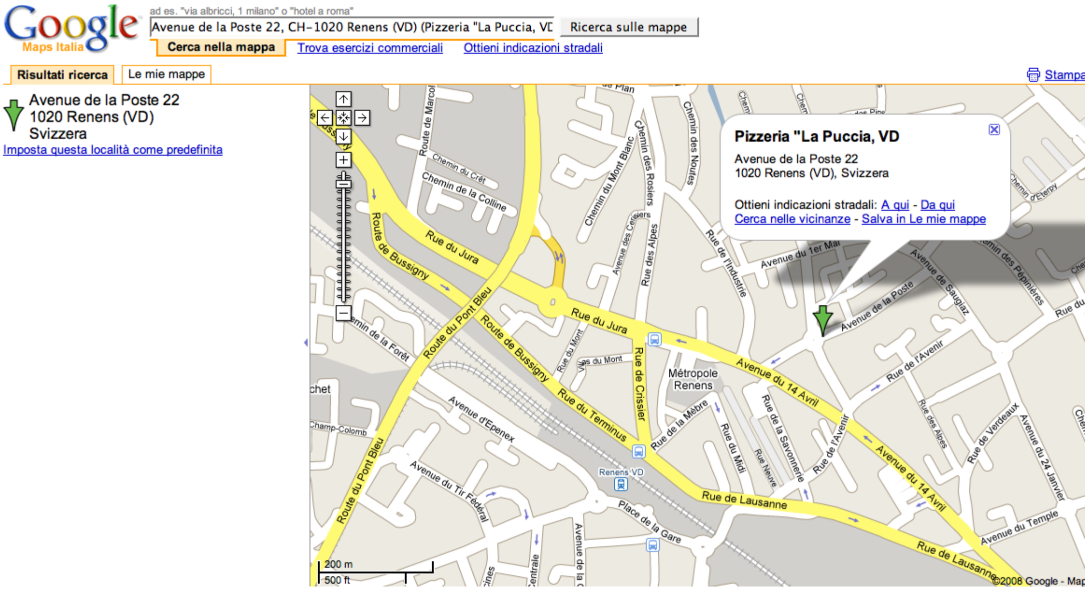


Figure 4-8: Screen-captures of ZoneTag. (Left) A picture was taken with the mobile's camera. (Center) Some tags are suggested based on the tags submitted by other users publishing content while being connected to the same cellular antenna. (Right) Eventually new tags can be suggested for the actual location



Avenue de la Poste 22, CH-1020 Renens (VD) (Pizzeria "La Puccia, VD

Avenue de la Poste 22
1020 Renens (VD)
Svizzera
Imposta questa località come predefinita

Pizzeria "La Puccia, VD
Avenue de la Poste 22
1020 Renens (VD), Svizzera

Ottieni indicazioni stradali: A qui - Da qui
Cerca nelle vicinanze - Salva in Le mie mappe

Figure 4-9: A screenshot of GoogleMaps. Annotations are displayed as bubbles over the geographical map

text and with geographical coordinates derived from the cellular network or input manually by the user.

However, other CAS applications developed for fixed computing environments use sophisticated geographical information systems and therefore can handle much more precise positioning information using latitude and longitude. GoogleMaps falls into this category of applications. The service was recently upgraded to allow users to create and share annotations with other users. Google provides maps with a resolution of 100 meters per 120 pixels[12]. By clicking on a point of these maps, it is possible to link short textual annotations to the geographical coordinates corresponding to the selected location. GoogleMaps presents only a few tools to manipulate the annotations. Linkage to the map, for instance, can only be punctual and not related to polygones (see figure 4-9). Services like GoogleMaps and Live! Maps were initially designed for being used on desktop computers but recently it has become possible to also use them on mobile phones. Each application allows for both synchronous and asynchronous interactions.

4.1.7 (g) Referent scope

CAS applications differ also in the portion of the map that is associated to the communicative content. A message might refer to a specific landmark (a point on the map) or to an entire building (a polygon on the map). The CAS tool might allow both kind of selection or more often it can allow a simple link between a point and the content. Ubiquitous CAS usually offer this latter selection approach, while more powerful fixed-use CAS might allow finer selection mechanism.

WikiMapia, described above, allows to select rectangles on the map, while iMapFan joins the SMS to the punctual position of the emitter of the message.

4.1.8 (h) Messages organization criterion

Most CAS applications organize their content by using the workspace to which content units are linked. Notes are displayed and retrieved using mainly a geographical criteria. In other words, messages that are displayed in close proximity probably describe similar features but they could have been produced by different users and at different times. Conversely, other applications may organize annotations according to their emission time, or else according to a particular discussion topic. The application ConcertChat described above falls in this second category, giving more importance to the conversation than to the geographical context.

A similar approach is implemented in UrbanTapestries[13], a CAS system for mobile devices developed by Proboscis, a non-profit company in England. UrbanTapestries is an experimental software platform for knowledge mapping and sharing, an activity that was defined by the cre-

[12] See `http://maps.google.com`, last retrieved January 2008.

[13] Proboscis, England Wales, UK. See `http://urbantapestries.net/`, last retrieved January 2008.

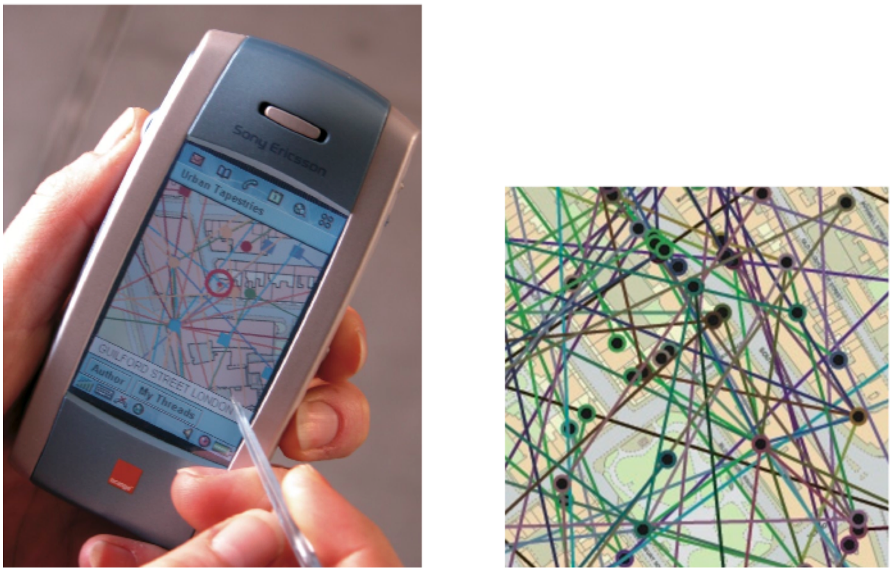


Figure 4-10: (Left) UrbanTapestries mobile interface. (Right) Pockets and threads around Bloomsbury Square, London (enlarged). These annotations were collected during the field trial of September 2004 (from Lane et al., 2005)

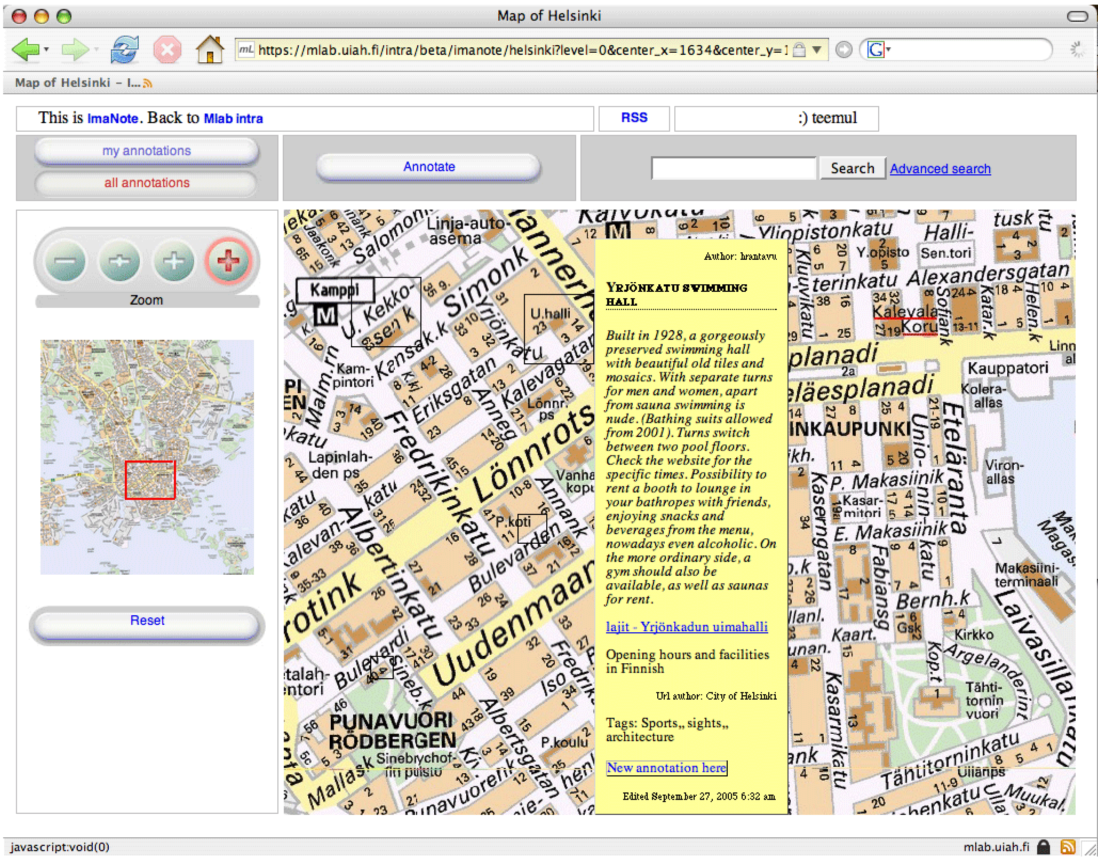
ators as *public authoring* (Lane et al., 2005). It combines mobile and internet technologies with geographic information systems to allow people to build relationships between places and to associate stories, information, pictures, sounds and videos with them. According to Lane, the goal of this application was to leave ephemeral traces of people's presence in the geography of the city:

> The Urban Tapestries software platform allows people to author their own virtual annotations of the city, enabling a community's collective memory to grow organically, allowing ordinary citizens to embed social knowledge in the new wireless landscape of the city. People can add new locations, location content and the 'threads' which link individual locations to local contexts, which are accessed via handheld devices such as PDAs and mobile phones.[14]

The organizing principle of UrbanTapestries was therefore the thread of messages, or *pockets*, in which users could discuss the same topic regardless of the location to which the different messages were attached. A message was visualized as a dot on the shared map, while a thread was visualized as a line connecting a series of dots (see figure 4-10). An example of a thread could have been the historical sites in a certain part of the city.

While applications like ConcertChat and UrbanTapestries give more importance to the flow of the conversation, ordering the messages by time and by thread, respectively, most of CAS

[14] These sentences were taken from the research web site, see `http://research.urbantapestries.net/`, last retrieved January 2008.



Author: hrantavu

YRJÖNKATU SWIMMING HALL

Built in 1928, a gorgeously preserved swimming hall with beautiful old tiles and mosaics. With separate turns for men and women, apart from sauna swimming is nude. (Bathing suits allowed from 2001). Turns switch between two pool floors. Check the website for the specific times. Possibility to rent a booth to lounge in your bathropes with friends, enjoying snacks and beverages from the menu, nowadays even alcoholic. On the more ordinary side, a gym should also be available, as well as saunas for rent.

lajit - Yrjönkadun uimahalli

Opening hours and facilities in Finnish

Url author: City of Helsinki

Tags: Sports,, sights,, architecture

New annotation here

Edited September 27, 2005 6:32 am

Figure 4-11: The interface of ImaNote. Notes are represented as yellow sticky notes overlaid on the map. Messages are anchored to squares on the map

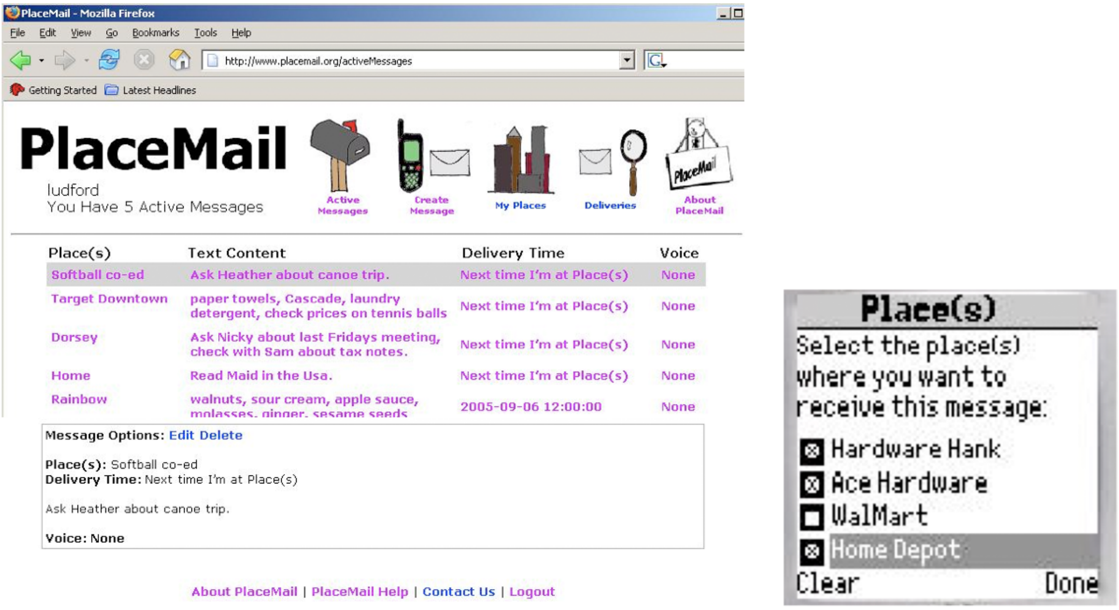
applications do the exact opposite, giving more importance to the spatial context to which the messages link. An example of this approach is ImaNote[15] (Image and Map Annotation Notebook), a web-based multi-user tool that allows users to display a high-resolution image or a collection of images online and add annotations and links in to them (Salgado & Diaz-Kommonen, 2006). The application creates square anchors on the image. Clicking on one of the anchors overlays the annotation on the image (see figure 4-11).

4.1.9 (i) Presentation of the referencing link

Different strategies exist for presenting the connection between the communication content and the context it refers to in the shared workspace. The interface of an application, for instance, can present the comunicative content visually as sticky notes attached to a map or the annotations can be listed on the side of the map and linked to specific locations with arrows or referring symbols (e.g., numbers or icons). Alternatively, a multi-media interaction mechanism can be

[15] ImaNote was developed at the Media Lab at the University of Art and Design Helsinki, Finland. See `http://imanote.uiah.fi/`, last retrieved January 2008.

used to deliver the communicative content: when the user of the application enters a specific geographical area, then the message gets delivered through a mobile device.


PlaceMail - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.placemail.org/activeMessages
Getting Started Latest Headlines
Active Messages
Create Message
My Places
Deliveries
About PlaceMail

PlaceMail

ludford
You Have 5 Active Messages

Place(s)	Text Content	Delivery Time	Voice
Softball co-ed	Ask Heather about canoe trip.	Next time I'm at Place(s)	None
Target Downtown	paper towels, Cascade, laundry detergent, check prices on tennis balls	Next time I'm at Place(s)	None
Dorsey	Ask Nicky about last Fridays meeting, check with Sam about tax notes.	Next time I'm at Place(s)	None
Home	Read Maid in the Usa.	Next time I'm at Place(s)	None
Rainbow	walnuts, sour cream, apple sauce, molasses, ginger, sesame seeds	2005-09-06 12:00:00	None

Message Options: Edit Delete

Place(s): Softball co-ed
Delivery Time: Next time I'm at Place(s)

Ask Heather about canoe trip.

Voice: None

About PlaceMail | PlaceMail Help | Contact Us | Logout

Place(s)

Select the place(s) where you want to receive this message:

☒ Hardware Hank
☒ Ace Hardware
□ WalMart
☒ Home Depot

Clear Done

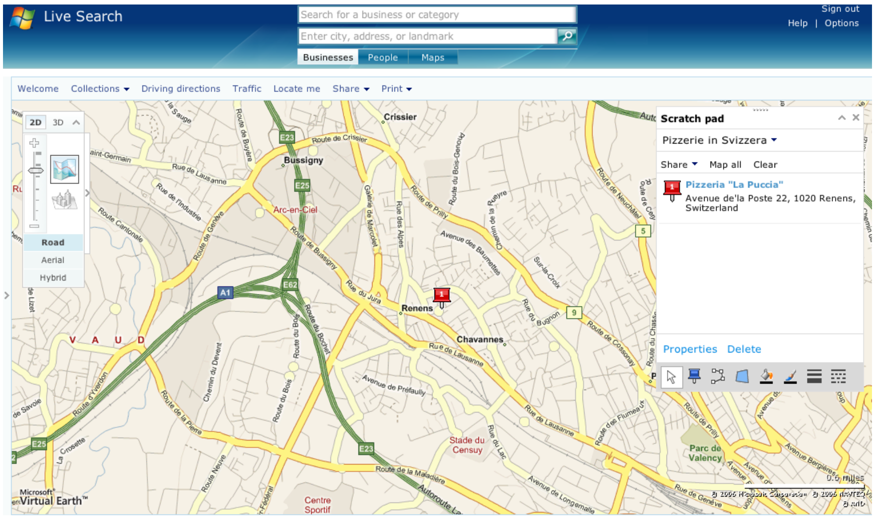
Figure 4-12: (Left) The PlaceMail web interface. (Right) The PlaceMail mobile interface using which the user can specify places for a message to be delivered (from Ludford et al., 2006)

An application following this last strategy is PlaceMail[16] (Ludford et al., 2006); Lundford, 2007). It allows users to define location-based reminders for their own use or that of their peers. A typical reminder is defined with a textual description and geo-temporal coordinates at which this description should be delivered to a mobile user. Figure 4-12 shows the web interface and the mobile interface of PlaceMail. A typical use of this application is to remind yourself to buy grocery items when passing near a shop on the way back from work.

On the contrary, when the referencing link is presented visually over a map, then messages can be overlaid at the referred locations or on a side, linked to landmarks. This is the case of Microsoft Live! Maps[17], and many other systems. Figure 4-13 shows a screen-capture of the Microsoft Live! Maps web site. A landmark on the map is numbered and the same number is then showed in a side text. Notes published on Live! Maps are organized in groups called *collections*. These can be either publicly shared or kept private. This service allows for rich interaction with the map. Messages, for instance, can be attached either to a point or to a user-selected polygon, like a building or a street.

Messages can be linked to the map using visible traits like arrows, as in the case of ConcertChat, a desktop application developed at Fraunhofer Institute of Integrated Information Systems (IPSI)

[16] See http://www.grouplens.org/, last retrieved January 2008.
[17] See http://intl.local.live.com/, last retrieved January 2008.

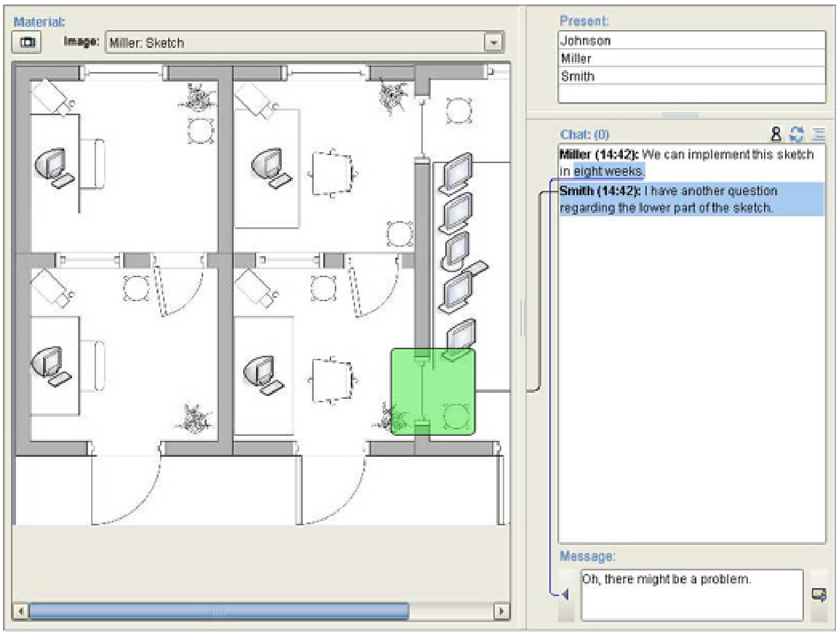


Pizzerie in Svizzera

Pizzeria "La Puccia"
Avenue de'la Poste 22, 1020 Renens, Switzerland

Figure 4-13: Annotations on Microsoft Live! Maps are linked symbolically to the side text

in Darmstadt, Germany[18]. ConcertChat is a standard chat application that allows the users to link a particular message to a rectangular area of the shared workspace. Once sent to the chat participants, an enriched message appears connected by an arrow to a green semi-transparent selection area (see figure 4-14). This application is designed for real-time discussions concerning a shared document (Mühlpfordt & Wessner, 2005).



Miller: Sketch

Johnson
Miller
Smith

Miller (14:42): We can implement this sketch in eight weeks.
Smith (14:42): I have another question regarding the lower part of the sketch.

Oh, there might be a problem.

Figure 4-14: ConcertChat interface. Messages linked to the shared workspace are connected by an arrow to a green semi-transparent selection rectangle

[18] See http://www.ipsi.fraunhofer.de/concert/, last retrieved January 2008.

4.1.10 Synthesis: CAS systems use style

This section presented applications that implement in different ways and for different purposes the mechanism of Explicit Referencing. These systems are summarized in table 4.1. Globally, CAS systems can be divided in two groups of applications: (1) mobile applications that aim at sustaining knowledge sharing and (2) fixed computing interfaces for synchronous collaboration. Systems in the first category are usually designed for an informal community use and they often support automatic detection of the user's location. Their interface offers a basic set of functionalities and messages are usually organized using the geographical context to which they link. On the contrary, systems in the second category are usually designed for small groups interacting in formal settings. Their interface presents a wider set of features.

In this last category of tools, links between the communication channel and the shared display can be made explicit using different mechanisms: (a) messages can be overlaid on the shared map, as in UrbanTapestries, iMapFan or ImaNote, for example, or (b) live in a separate space, linked to landmarks with symbols or traits, as in GoogleMaps, DisMoiOu, or ConcertChat.

Applications can differ not only in terms of how the explicit referencing link is created but also in how the messages are organized in the system, and more specifically whether more importance is given to the flow of conversation (e.g., ConcertChat) or to the geographical context to which the messages link (e.g., ImaNote).

As I will discuss in section 4.3, many of these systems have been developed in the academic milieu and have undergone a formal evaluation. Many others CAS systems have been developed by private companies that were interested in possible commercial exploitation of mobile maps. Finally, some projects involving CAS systems have been carried out by technology enthusiasts who were interested in the possible use of such technology for social applications and in connection with urban planning. I will discuss this last approach in the next section.

4.2 Collaborative annotations and urban planning

Recently researchers, artists, urban planners, and designers became interested in the practice of using portable electronic devices to annotate the urban landscape and sharing these notes publicly. This activity was defined as *placelogging* by Kottamasu (2007). Placelogging allows stories that happen at particular places and times to be indexed and accessed at the same geographical locations where they happened. Kottamasu conducted a comprehensive evaluation of many CAS systems. He described their narrative component, communicating experience of or reflection about places. Also, he described how the systems falls into one or more of six categories (p. 9):

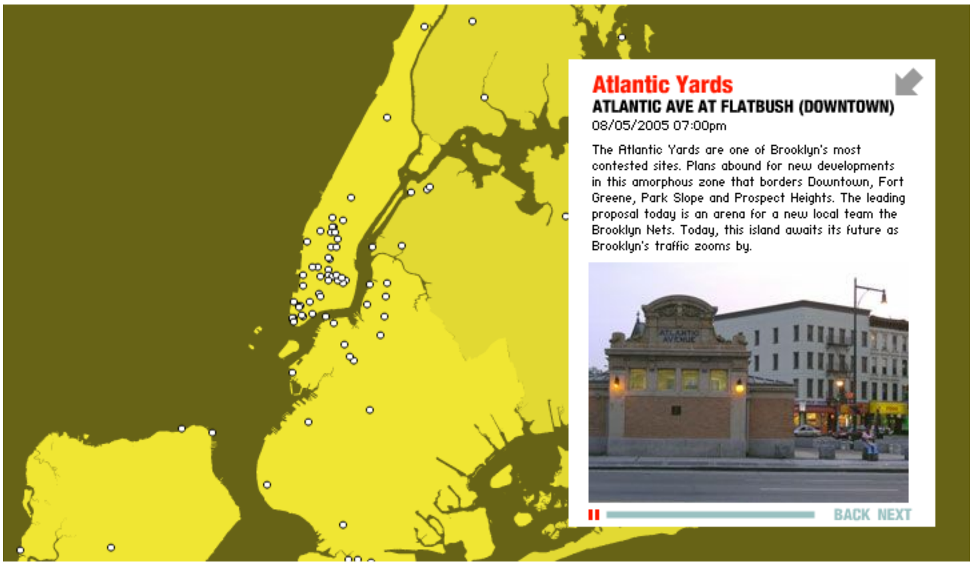
– *storytelling*: seeks to share personal thoughts or memories tied to places;

Table 4.1: CAS system classification according to the criteria discussed. This table resume all the applications discussed in this chapter. The first column indicates the section in which the specific application is described

CAS system	(a) group size	(b) interaction time-span	(c) geographical scale	(d) degree of immersion	(e) location acquisition [i] input; [r] retrieval	(f) referent description	(g) referent scope	(h) organization criterion	(i) presentation of referencing link
GeoConcept	Group	Asynchronous	City	Ubiquitous / Fixed	[i] Manual; [r] Manual	Coordinates	Point / Polygon	Map	Symbolic link
world66	Community	Asynchronous	World	Fixed	[i] Manual; [r] Manual	Coordinates	Point	Map	Overlaid text
iMapFan	Group	Synchronous	City	Ubiquitous	[i] Auto; [r] Auto	Coordinates	Point	Map	Overlaid text
dismoiou	Community	Asynchronous	World	Fixed	[i] Manual; [r] Manual	Coordinates	Point	Map	Overlaid text
FieldMap	Group	Asynchronous	Small area	Ubiquitous	[i] Auto; [r] Auto	Coordinates	Polygon	Map	Symbolic link
PlaceMail	Individual	Asynchronous	City	Ubiquitous / Fixed	[i] Manual; [r] Auto	Cell-ID	Point	Time / Map	Multimedia reminder
Live! Maps	Group / Community	Asynchronous	World	Fixed	[i] Auto; [r] Manual	Coordinates	Point / Polygon	Map	Symbolic link
ConcertChat	Group	Synchronous	Small area	Fixed	[i] Manual; [r] Manual	Coordinates	Polygon	Time	Arrow link
UrbanTapestries	Group / Community	Asynchronous	City	Ubiquitous / Fixed	[i] Manual; [r] Auto	Coordinates	Point / Polygon	Map / thread	Overlaid text and arrow link between treads
ImaNote	Group	Asynchronous	Small area	Fixed	[i] Manual; [r] Manual	Coordinates	Polygon	Map	Symbolic link
ZoneTag	Groups / Community	Asynchronous	World	Ubiquitous	[i] Auto; [r] Manual	Cell-ID	Point	GSM location	Symbolic link
GoogleMaps	Group / Community	Asynchronous	World	Fixed	[i] Manual; [r] Manual	Coordinates	Point	Map	Overlaid text
MapChat	Group	Synchronous	Small area	Fixed	[i] Manual; [r] Manual	Coordinates	Point / Polygon	Time	Overlaid text
WikiMapia	Community	Asynchronous	World	Fixed	[i] Manual; [r] Manual	Coordinates	Polygon	Map	Symbolic link
YellowArrow	Community	Asynchronous	World	Ubiquitous / Fixed	[i] Manual; [r] Code	Street name	Point	Location	Multimedia code
[murmur]	Community	Asynchronous	City	Ubiquitous / Fixed	[i] Manual; [r] Code	Coordinates	Point	Location	Multimedia code
GeoNotes	Community	Asynchronous	Small area	Ubiquitous	[i] Auto; [r] Auto	WiFi antenna / room	Point	WiFi Location	Multimedia reminder
ActiveCampus	Community	Synchronous	Small area	Ubiquitous	[i] Auto; [r] Auto	Coordinates	Point	Map	Symbolic link
E-Graffiti	Community	Asynchronous	Small area	Ubiquitous	[i] Auto; [r]Auto	WiFi antenna / room	Point	Map	Multimedia reminder
Location Linked Notes	Community	Asynchronous	Small area	Ubiquitous	[i] Auto; [r] Manual	WiFi antenna / Cell-ID	Point	Map	Multimedia reminder
Place-Its	Individual	Asynchronous	City	Ubiquitous	[i] Manual; [r]Auto	Cell-ID	Point	Location	Multimedia reminder
DeDe	Individual	Asynchronous	City	Ubiquitous	[i] Manual; [r]Auto	Cell-ID	Point	Location	Multimedia reminder
ShareScape	Community	Asynchronous	City	Fixed	[i] Manual; [r]Manual	Coordinates	Point	Map	Overlaid text

- *expression*: allows a channel for location-linked and mobile-accessible digital art, whether text, image, audio or video-based;
- *platform*: offers citizens an opportunity to voice and share opinions about stores, restaurants, public art, development projects, the state of neighborhoods;
- *guide*: provides information about available services, events and products at particular locations, for the benefit of both tourists and locals;
- *social network*: uses annotations as the basis for personal profiles through which users can identify shared interests and interact virtually;
- *document*: engages users in observing and recording the occurrence of noteworthy events or particular conditions of the surroundings.

As an example, Yellow Arrow[19] encourages users to leave 2"x3" stickers in places they wish to annotate. Each sticker bears a unique code, with which the user associates a multimedia message to a chosen location. The user can call a specific phone number and dial the code. Then she can upload the content to a remote server. The inputted content is therefore associated to the unique number. Other users that then stumble upon the sticker can dial Yellow Arrow, and use the code to retrieve the content. Eventually, they can engage in a conversation with the original author. The service web site acts as an aggregator, displaying interactive maps of the multimedia messages created by the users (see figure 4-15).



Atlantic Yards

ATLANTIC AVE AT FLATBUSH (DOWNTOWN)

08/05/2005 07:00pm

The Atlantic Yards are one of Brooklyn's most contested sites. Plans abound for new developments in this amorphous zone that borders Downtown, Fort Greene, Park Slope and Prospect Heights. The leading proposal today is an arena for a new local team the Brooklyn Nets. Today, this island awaits its future as Brooklyn's traffic zooms by.

Figure 4-15: A screenshot from YellowArrow.net. The user-created content is aggregated in interactive maps

[19] See `http://yellowarrow.net/`, last retrieved January 2008.

A similar project is [murmur][20]. It is an archival audio project that collects and curates stories set in specific locations in Toronto, Canada. A green sign, in the shape of an ear, with a telephone number and a location code printed on it, is placed on a telephone post or lamppost at each location where a story is available to cell phone users. Messages collected in the system can be only audio and can be accessible via cell phone to any passer-by. However, most stories are collected by staff onsite, while other users are allowed to add to existing stories with their anecdotes about that particular place. The stories are also available through the project website (see figure 4-16).

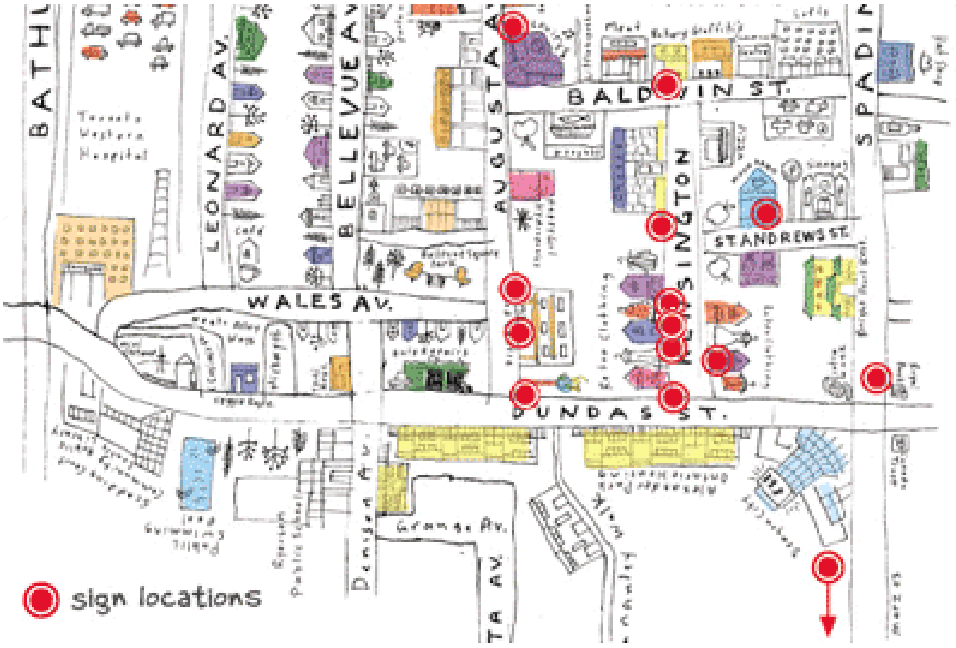


Figure 4-16: [murmur] interactive map of Kensington Market, in Toronto. Clicking on the red signs allows the user to retrieve the posted stories

As a last example, Moed proposed an interesting work on digital narratives (2002). She developed a system named DUMBO (Down Under Manhattan Bridge Overpass), an electronically delivered walking tour of the old Brooklyn Waterfront. The main idea of this project was that the user was delivered with audio information at specific locations. The user could also submit comments and reports about the places they encountered.

The reasons why these projects were of interests to urban planners is that in recent decades there has been a growing interest in incorporating public input and participation in decision-making processes affecting the development of the city. In this context, placelogging could assist in engaging community awareness and participation, as well as, making local knowledge visible and identifying the assets and gauging the pros and cons of proposed plans. Many projects

[20] See http://murmurtoronto.ca/, last retrieved January 2008.

in this area specifically position themselves in opposition to traditional planning practices. For instance, Jungnickel states (2004, p. 6):

> The possibilities for using these technologies to weave our own structures of narrative and creation through the fabric of the city enable a radical shift of capabilities, allowing for people to become both their own urban planners, defining their own visions of the city, or... designers of new conduits for navigating urban experience.

Kottamasu (*ibidem*, p. 17) suggests that subjective experience, organized by placelogging, can be a tool in rejecting and replacing the city as a given, an argument that Lynch also disseminated with his work (1964). Lynch argued that there is only a singular public image of the city based on the legibility[21] of its paths, nodes, districts, edges and landmarks. Lynch was aware that memory and meaning could contribute to the weight of the definition of these elements, however he argued that formal enhancements of the landscape alone could determine the city's identity. Currently, new technology allows citizens to take active roles in the definition and interpretation of the built environment. Meanings and memories are dynamically shared and created across many communities. Perhaps, Lynch's assumptions are now being challenged by how these new technologies help people form their image of the city.

4.3 Empirical studies of collaborative annotation systems

Over the last decade, there have been a growing number of projects on connecting information to geographical positions. Many of these were studied in a formal framework aiming at explaining the rationale of people using the system. This was the case for GeoNotes (Persson & Fagerberg, 2002), a prototype developed at SICS in Sweden, ActiveCampus (Griswold et al., 2003), developed at the University of California San Diego, E-graffiti, developed at Cornell University (Burrell & Gay, 2002), Location Linked Notes, developed at Virginia Tech (Tungare et al., 2006), and finally UrbanTapestries, developed by a non-profit company in the United Kingdom (Lane et al., 2005). All these prototypes allow users to connect opinions, preferences, recommendations, questions, and jokes to specific places.

The design rationale of GeoNotes (Persson & Fagerberg, 2002) was to endorse an open information space, where users could connect notes with spaces. Each note was then categorized according to the room corresponding to the position where it was generated. A specific design choice made it impossible to read/write notes from a remote position, as according to the authors, this would endanger the connection between the note and its spatial context would be endangered. The evaluation conducted with 80 students from a university community showed that in general people used the system for chatting with three main aims:

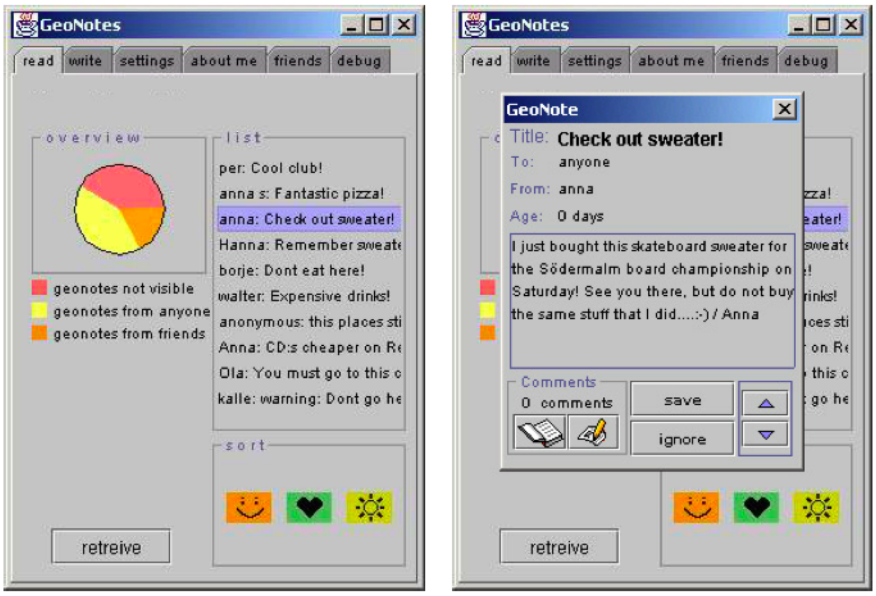
[21] The ease with which people understand the layout of a place.

1. *Object chat*, related to an object or physical aspect of the locale;

2. *Situation chat*, related to ongoing activities and situations in which several users took part;

3. *Talk-to-me chat*, an urge to chat with others independently of time and place.

The results showed that the triggers for authoring notes were not primarily physical objects or infrastructure, but rather the ongoing social activities and situations in that physical space (see figure 4-17).



per: Cool club!
anna s: Fantastic pizza!
anna: Check out sweater!
Hanna: Remember sweat
borje: Dont eat here!
walter: Expensive drinks!
anonymous: this places sti
Anna: CD:s cheaper on R
Ola: You must go to this c
kalle: warning: Dont go he

Title: Check out sweater!
To: anyone
From: anna
Age: 0 days

I just bought this skateboard sweater for the Södermalm board championship on Saturday! See you there, but do not buy the same stuff that I did....:-) / Anna

0 comments

Figure 4-17: GeoNotes interface for handheld. On the left-hand side the main screen with the list of the titles of individual GeoNotes. The small icons on the bottom of the list support sorting by notes directed to the user, the most popular notes, and the most recent notes, respectively. On the right-hand side, clicking on one of the title in the list brings up the individual GeoNote (from Persson & Fagerberg, 2002)

A similar setup was used in the ActiveCampus project (Griswold et al., 2003), where the researchers chose to create a viral community[22] because as they stated, for project sustainability they had to increase the application value. The authors argued that the social value of the application increases with the number of users. One of their most interesting findings concerned the analysis of the actual locations of the sender and receiver of a message. The application continuously logged the position of the participating students (the interface is represented in figure 4-18). The analysis showed that for 473 out of 539 logged pairs, the actual distance between the users when messaging was shorter than average distance of the same users during the rest of the day. Therefore they showed that participants were co-located while they were using the

[22] A community created leveraging on social networks. The term was adapted from viral marketing, an expression that refers to marketing techniques that use preexisting social networks to produce increases in brand awareness or to achieve other marketing objectives.

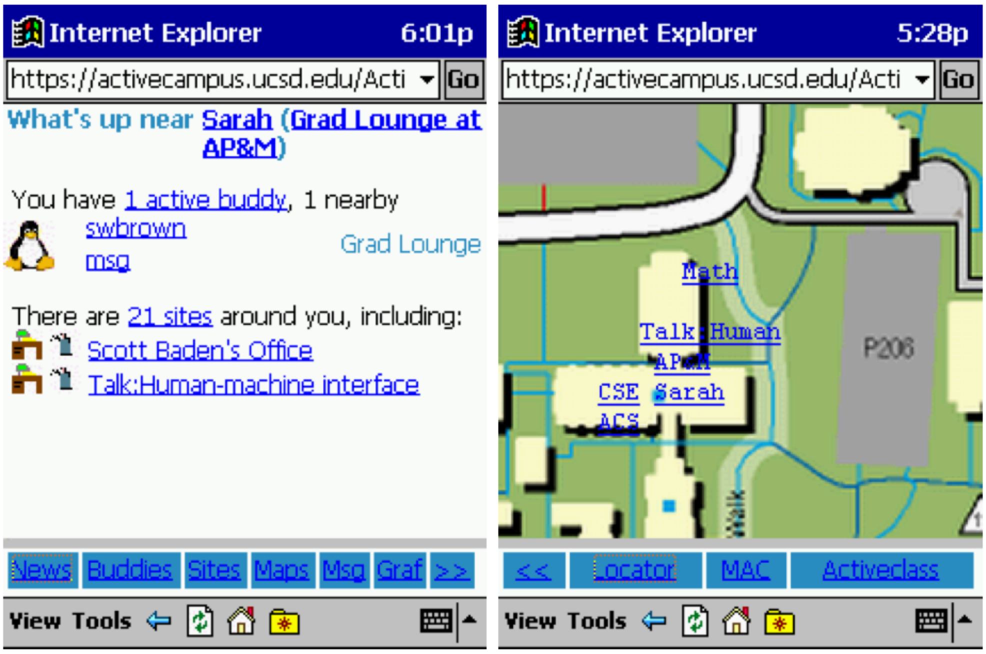
system to chat. In short, relative location as a context seems to matter in community-oriented computing.

The importance of a *critical mass* of messages/users was noticed in the evaluation of the E-Graffiti project (Burrell & Gay, 2002). The application interface is presented in figure 4-19. The lessons learned included difficulties with a misleading conceptual model. In short, the designers expected an asynchronous use of the tool, but students repurposed it mainly for synchronous communication, similar to that seen in GeoNotes. Authors also noticed a certain lack of use due to the reliance on explicit user input: "the fewer people using the system, the fewer notes people will contribute and the less value other people will get out of the system by reading those notes" (Burrell & Gay, 2002, p. 309). Finally, the authors highlighted the need for a highly relevant contextual focus: as this kind of technology is not part of daily life, users did not really think about information in terms of location, did not know what notes to write, and did not really have anything to share with others at a location.

Contrary to GeoNotes or E-Graffiti, Tungare and colleagues (2006) investigated the interaction design of Location Linked Notes, a CAS application where users were able to retrieve and edit messages from remote locations. They were interested in understanding whether users would prefer to be alerted of new messages (as in a push paradigm) as opposed to manually retrieving new content posted in the system (as in a pull paradigm). They designed a multiplatform architecture that could be accessed from mobile devices such as PDAs and mobile phones and from the internet (see figure 4-20). They tested the system with eight university students that commented positively the possibility of retrieving and editing notes at remote location. They also stressed the importance of authoring content while being in a place and being automatically alerted of content referring to a particular area when the user enters that area.

The four projects described above were developed in university campuses. In contrast, UrbanTapestries (Lane et al., 2005) was designed for use by the general population, aiming at sharing pointers about the city the participants were living in (the interface of UrbanTapestries was presented in figure 4-10). During two field trials conducted in 2004 and 2005, they discerned a series of general feelings and trends about the process and relevance of public authoring to everyday life. One of the key issues was the interaction time: people expressed their need for quick and simple interactions while on the move as opposed to a richer interaction when they are at work or at home. Participants saw the application as a new way of engaging in conversations about places. These conversations are fragmented and happen over time as well as in space.

While these studies carried out some evaluation of the use of geographical messaging systems in a real context, they were either limited to university campuses or, in the case of UrbanTapestries they lacked a detailed analysis of logs providing information on the mobile application and the context of its use.



What's up near Sarah (Grad Lounge at AP&M)

You have 1 active buddy, 1 nearby

swbrown Grad Lounge

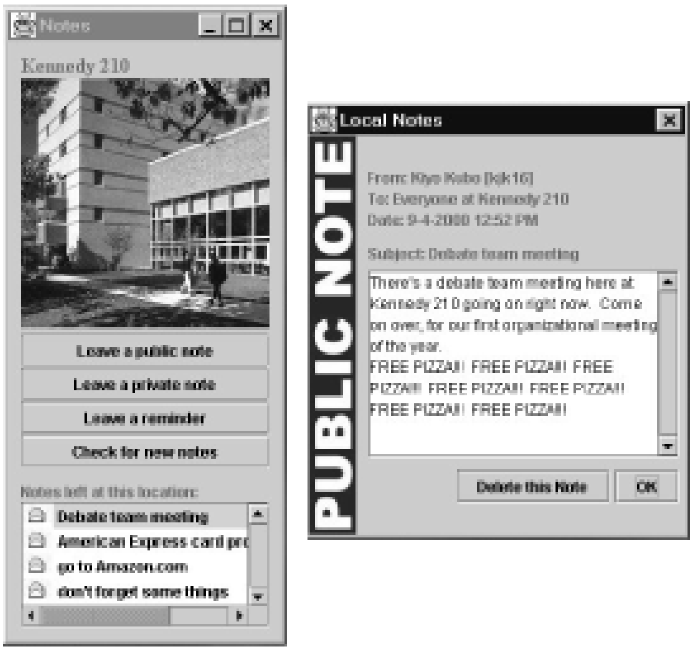
msg

There are 21 sites around you, including:

Scott Baden's Office

Talk:Human-machine interface

Figure 4-18: The *Map* and *Buddies* pages of ActiveCampus for a user "Sarah". "Maps" shows an outdoor or indoor map of the user's vicinity, with buddies, sites and activities overlaid as links at their location. "Buddies" shows colleagues and their locations, organized by their proximity. Icons to the left of a buddy's name depict the buddy on the map, begin a message to the buddy, and show the graffiti tagged on the buddy (from Griswold et al., 2003)



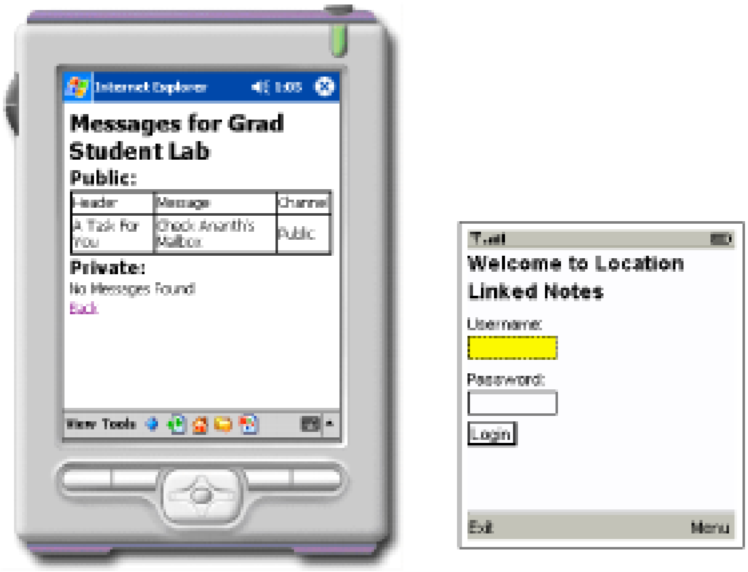
Notes left at this location:
- Debate team meeting
- American Express card pr
- go to Amazon.com
- don't forget some things

From: Kiyo Kobo [kjk16]
To: Everyone at Kennedy 210
Date: 9-4-2000 12:52 PM

Subject: Debate team meeting

There's a debate team meeting here at Kennedy 21 0 going on right now. Come on over, for our first organizational meeting of the year.
FREE PIZZA!! FREE PIZZA!! FREE PIZZA!! FREE PIZZA!! FREE PIZZA!! FREE PIZZA!! FREE PIZZA!!

Figure 4-19: Screen capture of E-Graffiti (from Burrell & Gay, 2002). The application tracked location using the WiFi network and it was designed for laptop users



Messages for Grad Student Lab

Public:

Header	Message	Channel
A Task For You	Check Ananth's Mailbox	Public

Private:

No Messages Found

Welcome to Location Linked Notes

Username:

Password:

Figure 4-20: Screen captures of Location Linked Notes system (from Tungare et al., 2006). (Right) Cellphone interface. Both screenshot have been captured using an emulator

Additionally, these systems were tested without a precise scenario or task in which the users were involved. These CAS applications were not designed around specific users' needs. Their purpose was to evaluate the generic idea that attaching virtual notes to physical location could be an attractive activity in which many people would spontaneously participate. Indeed, this was not the case as these projects registered a small level of participation during their trials.

Other researchers proposed the idea that these systems should have been connected more closely with a specific need like that for the user to be reminded of something while being in a specific location.

4.3.1 Location-Based Reminders

The idea of Location-Based Reminders, or LBR, is that a reminder is delivered for a location, namely the reminder is delivered near that location. This idea has been observed to be very useful as there is a common pattern for completing everyday tasks: people plan a certain action while being in a base (typically work or home), compose information resources like to-do lists and take these lists with them to refer to at the place where the task is performed (e.g., the grocery store).

Early proof-of-concepts of LBR include CybreMinder (Dey & Abowd, 2000), MemoClip (Beigl, 2000), and comMotion (Marmasse & Schmandt, 2000). These prototypes required ah-hoc hardware for location sensing detection that was not possible otherwise at that time. More recently researchers have implemented LBR systems that run on a cell phone.

Place-Its (Sohn et al., 2005), for instance, is one such application. The three components of a Place-It note are the trigger, the text and the place. The trigger defines whether the message should be signaled upon arrival to or departure of the associated place. The text is the message associated with the reminder and the place is the location defined by the user where the reminder should activate (see figure 4-21). Location sensing was achieved with PlaceLab (I. Smith et al., 2005). The authors conducted a user study with 10 subjects during a period of two weeks. They interviewed the subjects before and after the experiments. One interesting finding was the unexpected presence of 'motivators' reminders, a kind of reminder used to motivate a person to perform a certain task while being in a certain place. Participants who worked to a set time schedule mapped their relevant context to time cues and modified their behavior. Results showed that location was widely used as a cue for other contextual information. It appeared that the convenience and ubiquity of location-sensing provided outweighed some of the current weaknesses of the system.

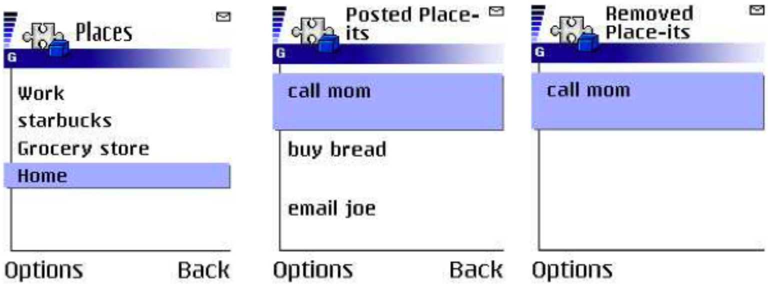


Figure 4-21: Some steps for creating a Place-Its note (from Sohn et al., 2005)

Jung and collaborators (2005) followed the same basic idea in the development of a context-enhanced mobile messaging system. They described DeDe, of which the central feature is to offer a definition of the delivery context for multimedia messages. Their starting assumption was that people need better support for conversational timing. They built an application that enables the user to define the delivery context of a certain message. The context was described either as a certain time of the day or a certain location (determined by the user and tracked with the Cell-ID). Additionally the user could define the context as the proxemics to another user or the activity of making a phone call to a certain user. They evaluated the system with a field test. The results showed that the DeDe system was used only when the user could predict in advance the message context of delivery. They based this inference on the knowledge of the receiver's schedule and her movements in the city. The main concern expressed by the subject involved in the test concerned the reliability of the delivery of the messages.

While Place-Its centered on opportunistic reminding, PlaceMail was developed to investigate current practices for managing personal everyday tasks (Ludford et al., 2006). PlaceMail allowed the user to send a message to himself/herself but instead of being directed towards an email client, it was received on a cell phone at a time and place specified by the user (the interface of PlaceMail was presented in figure 4-12). The system was tested during two weeks of field trial by 20 participants, who created 344 messages, with an average of 17 per person. Results showed that the traditional *geofence*[23] radius that is used to trigger the delivery of a message depends on people's patterns through an area and the geographic layout of the space. In particular, survey responses indicated inconclusive answers as to what the ideal distance is for delivering location-based information. The messages collected in the PlaceMail study served for a follow-up study on a possible application of location-based annotations, as I will discuss below.

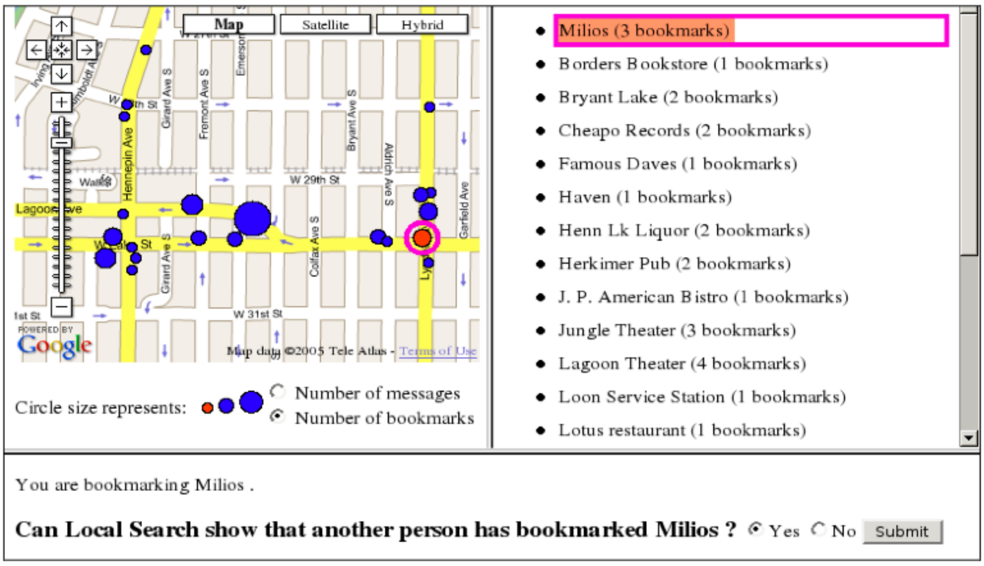
4.3.2 Application of Location-Based Annotations, some examples

Ludford and colleagues proposed a system for reusing notes that were generated for private needs to satisfy public activities (Ludford et al., 2007). They reused the messages collected during the field trial of PlaceMail to determine whether the authors of those annotations were willing to share their notes with strangers. Collected reminders were categorized in about 20 kinds of different places. They interviewed the authors of the messages asking whether they wanted other people to see their reminders. Results showed that consistently, participants dissented in showing reminders pointing to residential places and those containing people's information to public display. The authors used a different interface called ShareScape to collect local recommendations. They also seeded this database with data from the previous study. They tested the collected *placemarks* with a second group of testers who could run local searches to find place-related information (see figure 4-22). Users commented positively the cumulative ShareScape maps, saying that this map helped them for opportunistic planning of errands. This work also stimulated in-depth research on privacy issues that this sharing of information might have consequences for (Ludford, 2006).

A different application idea was pursued by Lemmelä and Korhonen (2007), who studied visualization techniques for datasets of location-based postings. They argued that current visualization and access methods do not scale well for a high number of geo-referenced messages. They proposed a semi-transparent heat map superimposed on the map workspace to visualize posting density in the area of interest. Additionally, they implemented automatic keywords extraction for supporting the search and the filtering of the posted messages. Their user evaluation reported positive results.

Location-based annotations might be used to generate anamorphosis maps for building visualizations that are useful for making comparisons (J. Lévy, 2005). These maps are also called density-equalising maps. Usually maps represent countries or regions according to their land area. An anamorphosis map re-sizes each country (or other geographical unit) according to some

[23] A circular boundary around a certain location that triggers an interaction mechanism.



- Milios (3 bookmarks)
- Borders Bookstore (1 bookmarks)
- Bryant Lake (2 bookmarks)
- Cheapo Records (2 bookmarks)
- Famous Daves (1 bookmarks)
- Haven (1 bookmarks)
- Henn Lk Liquor (2 bookmarks)
- Herkimer Pub (2 bookmarks)
- J. P. American Bistro (1 bookmarks)
- Jungle Theater (3 bookmarks)
- Lagoon Theater (4 bookmarks)
- Loon Service Station (1 bookmarks)
- Lotus restaurant (1 bookmarks)

Circle size represents: ● ● ● ○ Number of messages ● Number of bookmarks

You are bookmarking Milios .

Can Local Search show that another person has bookmarked Milios ? ● Yes ○ No Submit

Figure 4-22: The ShareScape interface (from Ludford et al., 2007)

other variable, for example population, number of people with AIDS, etc (see for instance the map[24] in figure 4-23).

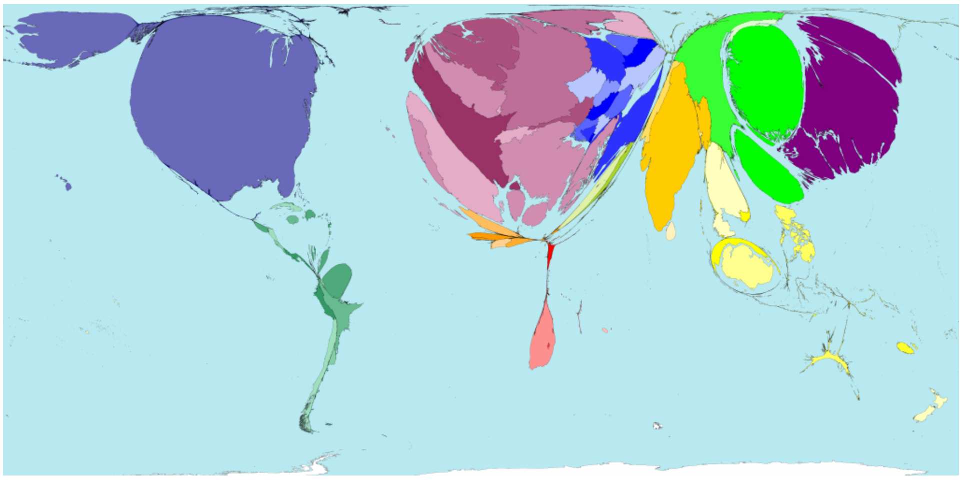


Figure 4-23: Anamorphosis map of crude petroleum import. Copyright SASI Group (University of Sheffield) and Mark Newman (University of Michigan)

As a final example, location-based annotation can be used to understand what people around us think. Jones and colleagues (Jones et al., 2007) devised a system to capture Google queries generate by their university community and displayed them on a map for other users to look at. Their experiences suggest that presenting users with other people's in situ queries influences their information seeking interactions positively.

[24] See `http://www.sasi.group.shef.ac.uk/worldmapper/`, last retrieved January 2008.

4.3.3 Synthesis: open questions

- *What are the reasons for annotating?* The studies discussed in this section reported inconclusive results on why people might annotate space. Many studies have been conducted under the assumption that connecting communication messages to the actual space through technology could be an interesting activity that people might want to pursue for the sake of sharing pointers with their community. However, the reported studies registered a low level of participation. On the other hand, systems which were designed with more precise scenarios in mind, like PlaceMail, registered a higher level of user participation and satisfaction. Therefore, it is difficult to draw strong conclusions about why people might want to annotate space. The studies discussed seem to suggest that practical reasons like that of being reminded to do something on a certain location might respond better to users' needs that the idea of contributing to the social navigation of a community.

- *Does actual location matter?* Many CAS designs forced the user to retrieve and publish content while being in the actual location to which the messages refer. However, results from GeoNotes showed that students-participants had the tendency to use the system to communicate about social activities rather than physical objects or infrastructure. Additionally, many messages were tagged to rooms which did not exist (e.g., "7th and a half floor") as they did not refer to a particular place. Similarly in ActiveCampus, participants posted messages when they were close to each other, similarly to what Japanese teenagers do with SMSs to enhance their co-located encounters (see section 2.3.4). So the question is whether location plays a role in these systems and particularly in the production of the messages. Studies reported in this section did not offer a conclusive answer.

- *What are the best rules of engagement? Do we have to provide a scenario of use?* Many system evaluations reported the difficulty in engaging participants to contribute to an empty system. The value of a CAS application increases with the number of annotations posted. The designers of E-Graffiti highlighted how these systems need a precise scenario of use as this kind of technology is not part of common practice and people are not used to compose messages according to the location to which they refer, even if this binding of language and space is natural. On the other hand, other evaluations, like that of Urban Tapestries showed a spontaneous adoption of the technology by its users. Therefore, the balance between scripting or designing flexible applications for annotating space is not clear.

- *What is the best way to protect CAS users' privacy?* One of the biggest challenges for every location-based application is to define solution to protect the privacy of the users. Other scholars have tackled the issue and suggested possible solutions (see for instance the work of Ludford and colleagues (Ludford et al., 2007)). Although of great importance, this thesis does not deal directly with this question.

4.4 Conclusion

This chapter presented different implementations of CAS tools that have been designed both for fixed workstations and for mobile devices. I proposed a classification based on nine criteria such as *the group size* for which the service is designed, *the way the referencing link is presented*, *the organization criterion*, and the *time span of the interaction*. Most of the research on CAS concentrated on designs for asynchronous and mobile use. Particularly, research on these systems was not conducted under precise research questions on how this technology could support collaborative work at a distance. Researchers studying these systems concentrated on the reasons and the conditions that could stimulate people to annotate, underestimating the importance of the collaborative dimension of this activity. Moreover, these studies reported inconclusive results on the conditions under which these applications might serve useful and practical purposes and the best parameters to deploy them.

Research concerning CAS lacks a systematic evaluation of the users' interactions with a system. The studies presented in this chapter either lack *details of analysis* (they did not have precise users' logs recording action in the system, e.g., E-Graffiti), *scale of annotation* (the focused on a small geographical area such a campus and an homogeneous community, e.g., ActiveCampus), or a proper *experimental evaluation* (e.g., Urban Tapestries). Additionally, the datasets produced during the field trial of these projects were not publicly available for supporting further research. Relevant details for a proper evaluation might include the position of the user while interacting with the system, a systematic recording of the actions performed in the system with actions such as running queries, applying filters, writing posts, and retrieving posts. The scale of annotation is also important for defining relevant annotations: notes posted by students on a campus might be biased towards social activities compared to notes directed to a general public and posted in a city-wide environment. Finally, some applications were developed for commercial or artistic purposes and did not undergo a systematic experimental evaluation. Therefore their results are not really comparable with more systematic studies.

Finally, the literature review presented in the previous chapters conjectured that CAS applications could improve collaboration both in terms of group cognitive processes, by ameliorating referential communication, and by enabling proper conditions of collaborations (e.g., by allowing affordances for communication or through the lessening of the effort required to ground contributions). The results achieved in this area are unclear. Furthermore, CAS applications and Explicit Referencing received little attention, as such. The research presented did not really deal with collaborative task performance, and cognitive processes (such as division of labor, interaction negotiation, and communication strategies). Moreover, when these interaction mechanisms have been studied, they were rarely the core research question but more part of the side results. Perhaps is time for a more systematic evaluation of CAS, one of the objectives of this thesis.

Chapter 5

Research Methodology

This chapter presents the research methodology that I used to design and analyze the experiments described in this thesis. It refines the scope of research on Collaborative Annotation Systems and the research question that I seek to answer as well as my methodological choices.

5.1 Research problem

The core of this work is to understand how spatial representations can be used to sustain human communication and interaction at a distance. The second chapter discussed how space and language are fundamentally intertwined: spatial context is used to make utterances more effective in conveying the intended meaning. Spatial situations are also schematized in the mind. They form cognitive maps that themselves have spatial features that influence how people think. Maps therefore condense this interplay of language and space, encoding spatial features in graphical elements. Maps are used in conjunction with language.

The third and the fourth chapter showed that in the field of CSCW and HCI many prototypes have been developed to take advantage of an explicit binding of the messages exchanged by remote collaborators, and the spatial elements of their shared workspace. Particularly, chapter 3 argued that providing deictic gestures support for conversants who are not co-located might improve collaboration, but also that different design choices concerning how to implement this mechanism might profoundly impact performance. Additionally, chapter 4 described Collaborative Annotation Systems in which a specific design was employed to support the binding of textual messages to a shared map. Again, the analysis of the available systems revealed the breadth of the possible design choices under which these systems can be implemented. Little research has been done to specifically address the design of CAS and the benefits of different interface mechanisms.

Specifically, CAS research lacks a systematic evaluation of the users' interaction with the system. The studies concerning ubiquitous CAS systems reported in the literature review either lack details of analysis, or lack a proper experimental evaluation. Little is known about the usage patterns of these applications, such as the way content is posted and retrieved, and the topics that are useful to discuss over a geographical support, and for what reasons. Finally, it would be interesting to understand whether using a CAS system under a specific scenario might change the way it is used. It seems pertinent to pose specific questions on the linguistic benefits of using CAS applications.

According to the vocabulary defined in the research framework presented in chapter 2, Explicit Referencing (ER) is a specific type of coordination device. This signal can be used by team members to share perspectives of contextual elements. ER can be produced and perceived through ad-hoc interface mechanisms and contributes to forming the participants' common ground. The use of coordination devices of this kind can help several aspects of the interaction: first, they can reduce misunderstandings, diminishing the range of possible interpretations that the receiver of a communication can assign to the message. Also, as ER-messages bind multi-modalities of communication, the linguistic channel is offloaded: messages pointing to specific points in space should require less effort to be produced. One last point might concern the interaction of ER with gaze. As chapter 3 discussed, gaze is used to monitor gestures and communicate attention, we might expect the use of explicit referencing to have an influence on the way the shared workspace is looked at by participants.

As one can imagine, different intents and usages can be expected from the use of a CAS system in a ubiquitous setting and from a fixed workstation. While in the former case, coordination of the user might comprehend actual positions and moves across a space, this might be less important in the latter case. Additionally, while a fixed-use CAS might invite synchronous interactions, CAS are more often designed and used for asynchronous communication (chapter 4 reported exceptions to this point, like the GeoNotes application). Therefore, this dimension appears relevant for the development of the research reported in this thesis.

5.1.1 Research questions

The basic goal of this thesis is to explore how technologies can improve remote collaborations taking advantage of a multi-modal combination of communication and spatial elements. The discussion in chapter 3 shows that the design of a mechanism for Explicit Referencing, particularly the way in which messages are linked to the shared workspace and organizing principles of the discussion, affects collaboration practices and performance. To fully evaluate this hypothesis several key questions must be addressed.

Previous research on Collaborative Annotation Systems demonstrated that messages enriched with spatial information might be used efficiently to support various human activities. However, it is not properly understood what specific applications CAS would support and the reasons and the modalities under which people might want to create location-based annotations. Particularly, what is not entirely clear is the value of mobility in the production of location-based annotations. As well, the role of a virtual map in supporting the production and retrieval of messages relating to actual locations.

Understanding how annotations are produced in ubiquitous CAS and analyzing the interplay of messages, map, physical locations, and communication will help assess how the design of these systems impacts their use and potential deployment. Therefore, I framed my first research question as: **Q1, what kind of messages do users associate to map locations?** Similarly, I am interested in understanding the relation of the produced messages with the locations to which these messages have been attached, and finally the role of the virtual map in supporting this link. The second research question I address is: **Q2, what is the relationship between the messages and the actual locations to which these messages refer?**. The third research question completes this triangular relation between the map, the message and the physical location to which the message refer (see figure 5-1): **Q3, how does the map mediate this relationship?** Finally, we ask whether people using a CAS in a precise task (e.g., producing notes required for their professional activities) might employ practices and strategies that are not directly comparable to those of another group of users that did not follow such precise task. Therefore, I framed my fourth research question as: **Q4, do the results of the first three research questions change when the participants use the CAS tool in a structured task?** Answering these questions requires the aggregation of user generated data from a real context of use of a ubiquitous CAS. Users' interaction with this application were tracked with a high level of details in order to address these questions. From the analysis of the applications reviewed in chapter 4, it is clear that the studies that have been conducted so far, lack the right level of detail in the logs of the tested applications. Also in the majority of cases, the produced datasets are not accessible for further research.

A more quantitative dimension of this research is concerned with how the provision of remote deixis, as a general interface mechanisms, might improve performance in collaborative tasks. Previous research discussed in chapter 3 demonstrated evidence that the provision of remote gesturing improves performance in collaboration. However, none of the reported studies actually focused on the specific mechanism of remote deixis. Remote gestures and the more specific telepointer solution cannot be assimilated to the process of specifically relating communication to the shared workspace, as in the case of remote deixis. While we can hypothesize that remote deixis improves collaboration performance, we need to understand how and why this occurs. Therefore my fifth research question was framed as: **Q5, does the availability of Explicit Referencing enhance the performance in a collaborative problem solving task at a distance?** The hypoth-

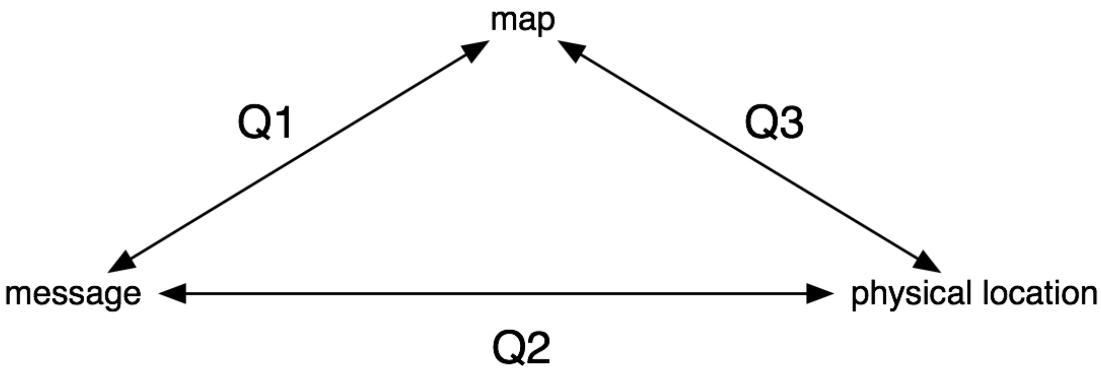


Figure 5-1: The initial three research questions and the way they inquiry the triangular relation between the map, the message, and the physical location to which the message refer

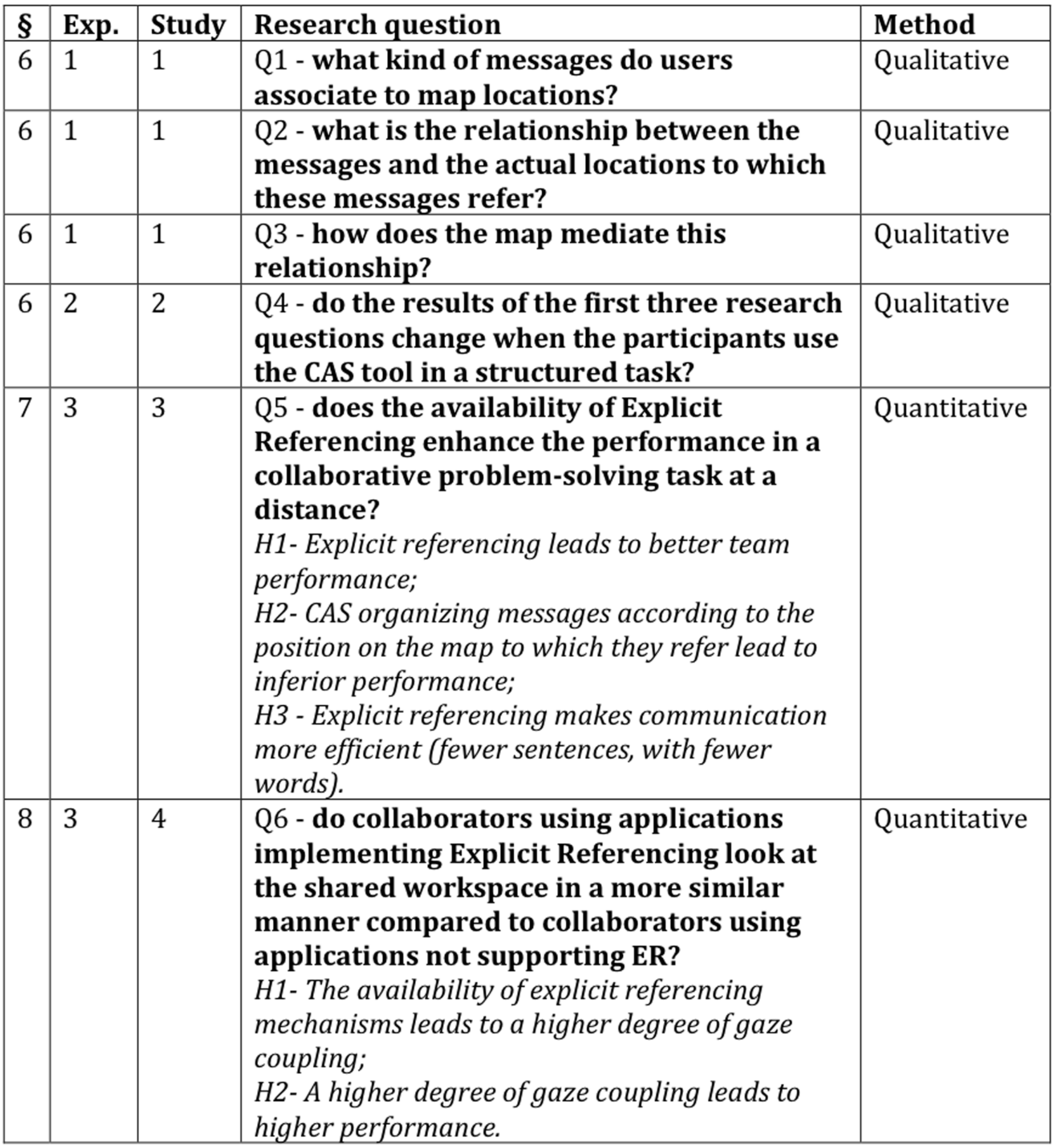
esis suggested by the literature is that making this information explicit helps in disambiguating references to shared objects, thus improving the collaboration process. However, if utterances are overlaid on a map, they are no longer sequentially displayed as in a chat window but appear scattered over the map. This dispersion of utterances may actually be detrimental to the joint maintenance of the context of the conversation. Therefore, my hypothesis is that the disruption of the conversation linearity has negative consequences for collaboration performance.

An additional issue to address is the relation of Explicit Referencing to gaze. From the literature, it is known that gaze is used in face-to-face communication to monitor the production of gestures but also as a communication device. At a distance, the ecology of these communication mechanisms is disembodied. However, gaze can still interact with the way information is presented in the remote display. Therefore, I consider relevant to pose the following question: **Q6, do collaborators using applications implementing Explicit Referencing look at the shared workspace in a more similar manner compared to collaborators using applications not supporting ER?** The hypothesis suggested by the literature is that the availability of shared reference points on the workspace should attract the attention of the participants. Of course eye-tracking technology was required to answer this question. Table 5.1 summarizes the research questions presented in this section.

5.1.2 Research scope

This thesis explores how CAS influence the cognitive processes occurring in collaboration. It consider collaborators and the artifacts used in their interaction as forming one global cognitive system. Therefore, I does not only focus on individual cognitive processes, like memory, and problem solving, but also on the mechanisms of group behavior (e.g., communication). In this philosophy, *cognitive processes* are the faculty of individuals to manage information. More

Table 5.1: Summary of general research questions



§	Exp.	Study	Research question	Method
6	1	1	Q1 - **what kind of messages do users associate to map locations?**	Qualitative
6	1	1	Q2 - **what is the relationship between the messages and the actual locations to which these messages refer?**	Qualitative
6	1	1	Q3 - **how does the map mediate this relationship?**	Qualitative
6	2	2	Q4 - **do the results of the first three research questions change when the participants use the CAS tool in a structured task?**	Qualitative
7	3	3	Q5 - **does the availability of Explicit Referencing enhance the performance in a collaborative problem-solving task at a distance?** *H1- Explicit referencing leads to better team performance;* *H2- CAS organizing messages according to the position on the map to which they refer lead to inferior performance;* *H3 - Explicit referencing makes communication more efficient (fewer sentences, with fewer words).*	Quantitative
8	3	4	Q6 - **do collaborators using applications implementing Explicit Referencing look at the shared workspace in a more similar manner compared to collaborators using applications not supporting ER?** *H1- The availability of explicit referencing mechanisms leads to a higher degree of gaze coupling;* *H2- A higher degree of gaze coupling leads to higher performance.*	Quantitative

specifically, this thesis will analyze cognitive processes that occur within groups. These enable the achievement of group performance and are the core components of group collaboration (Dillenbourg, 1999).

The framework introduced in section 2.4, at page 40, demonstrates this analytical approach. The experiments reported in the forthcoming chapters will analyze communication within groups through a psycholinguistic perspective. The location to which a message is attached acts as a coordination devices because it provide participants with a shared perspetive on contextual elements.

5.2 Methodological choices

This thesis presents four studies that have been designed to answer the research questions presented in the previous section. These studies correspond to three experiments that have been conducted in two different technological contexts. The first and the second experiment have been carried out using an ubiquitous CAS and that was tested in a field study. The third experiment was conducted indoor using desktop CAS applications. The formers are preliminary experiments that helped me shape the third experiment, which has a broader scope and that was used in three different studies. Table 5.1, at page 143, summarizes the different studies conducted and the research questions they address. Strictly speaking, the studies conducted in a physical environment are not comparable with the lab studies, but I aimed rather at understanding the salient and common trends in both environments.

The first and second were designed by following an observational approach to research where the aim was to extensively document the practices, customs and interactions of the ubiquitous CAS users as they are performed in their natural environment (Robson, 2002). However, as one of the goal was to record precise and objective logs of the users' interaction with the system, a *field experiment* approach (Goodman et al., 2004) was adopted. Field experiments are quantitative evaluations conducted in the field, drawing from aspects of both qualitative field studies and lab experiments. In applying this approach, the emphasis was put on the analysis of conversational behaviour to derive the impact on the use of technology had on conversational exchanges. In fact, these location-based applications support brand new experiences. A possible approach was proposed by Crabtree (2004): it consists in deploying new technologies in the wild and treating them as *breaching experiments*, allowing to provoke practices and reveal contingencies between activities and technological interactions.

The three last studies, on the other hand, refer to a controlled experiment designed to strictly compare different implementations of CAS applications. In this analysis, Attention was taken in reducing the effects of confounding variables such that key independent variables could be manipulated and effects on dependent variables therefore measured (see next section). As it will be explained in chapter 7, the availability of two factors was manipulated, namely an Explicit Referencing mechanism and a linear message history. Particularly, I chose to use existing CAS applications to maintain a degree of ecological validity. As a result, I employed applications that differed slightly in the sets of features available to the users. This choice was inspired from the notion of quasi-experiment developed by Cook and Campbell (1979).

To conclude, there are acknowledged strengths and weaknesses to both quantitative and qualitative approaches. The reasons why these approaches were chosen for the three experiments lies in the research question that I aimed to answer. It is apparent that an observational approach would best suit the situated analysis of collaboration in a social context, in which the first four research questions should be answered. However, providing responses to the other questions necessitates the comparisons of conditions and a fine-grained control of the interfering variables that is best accomplished through a controlled experiment.

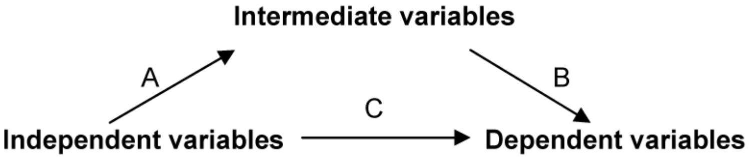


Figure 5-2: The interactions paradigm as explained by Dillenbourg et al. (1996). In this thesis the process variables are represented by linguistic markers like number of words or quality of the turn tacking and by gaze markers

5.2.1 Rationale for the controlled experiment

The controlled experiment presented in this work focuses on collaborative settings. It is conducted according to the *interactions paradigm*, and dissociates three types of questions (see figure 5-2):

a) What is the effect of the experimental conditions (e.g., presence or absence of ER) on intermediate variables (e.g., word frequency)?

b) What is the relationship between categories of interaction and collaboration performance, i.e., the effects these interactions may have on collaboration score or the time required to complete the task?

c) How do the intermediate variables (e.g., word frequency) mediate the relationship between the experimental conditions and collaboration outcomes?

The mediation effect of the intermediate variable will be assessed using the technique developed by Baron and Kenny (1986). In their vocabulary, path C of figure 5-2 is called the *direct effect*.

The mediator has been called an intervening or process variable. Complete mediation is the case in which an independent variable no longer affects a dependent variable after an intermediate variable has been controlled and so effect C is zero. Partial mediation is the case in which the path from independent variable to dependent variable is reduced in absolute size but is still different from zero when the mediator is controlled.

The eye-tracking analysis will be conducted under a slightly different paradigm. The relationship between gaze patterns and collaborative processes can be tackled at two levels of granularity. At a low granularity level (chapter 8), I will investigate whether global gaze parameters are related to the quality of collaboration or collaboration performance. At a higher granularity level (chapter 9), the social interaction log files (dialogues and actions in the shared space) will be segmented into critical episodes, known to contribute to collaboration. The set of gaze paths occurring during these episodes is then analyzed by a supervised algorithm to infer which gaze patterns predict these interaction patterns. The opposite analysis will be also performed: an algorithm segment the whole gaze path into episodes that will be interpreted by the experimenter by analyzing social interactions that occurred at the same time.

5.2.2 Unit of analysis

In the analysis of the quantitative experiment, I will focus on group cognitive processes. It seems obvious that the unit of analysis is the group. However, Kenny et al. (1998) described how the non-independence of the observations could be problematic. If the individual is used as unit of analysis, then the assumptions of independence are likely to be violated as group members might influence one another. Alternatively, if groups are used, the power of statistical tests are likely to be reduced, because there are fewer degrees of freedom than in an analysis which uses individuals. A possible solution is to use multi-level analysis since it allows more flexibility. Alternatively, Kenny (1998) explains that a simpler method of measuring the non-independence of the data is the use of the *intraclass correlation*. These statistics can be viewed as the amount of variance in the persons' score that is due to the group, controlling the effects of the variable.

In the last study reported in chapter 9, I conducted a micro-analysis of the movements of gaze associated with a particular utterance. In this case, my unit of analysis were the single utterances. In this case, I had the opposite problem as results stemmed from utterances. A participant produced many utterances, therefore possibly biasing the results. In these situations, I computed intraclass correlations to determine the non-independence of the measures. However, for certain group measures like performance, the group was necessarily the unit of analysis as the measures referred to joint activity that could not be carried out by individuals alone.

5.3 Custom-made CAS tools

Answering the questions of this thesis required the implementation of prototype CAS tools that I could strictly control and that could help assess differences in design features compared to existing applications. For the first and second experiment, an ubiquitous CAS was built, named STAMPS, that allowed to log every user's action that was performed while the user was connected to the system[1]. The third experiment compared desktop CAS tools that differed in the way messages were associated to a shared map. Therefore, a desktop application supporting synchronous communication was needed, which organized messages according to the geographical context of the reference, instead of the emission time, and for which messages' anchors were superimposed on the map. As I could not find an application with these requirements and that I could easily experiment with, a second CAS tool for a desktop computer was built, called ShoutSpace[2]. These two systems wil be described next.

5.3.1 STAMPS

STAMPS is an application for Symbian Series 60 smartphones. I chose a mobile phone instead of a Personal Digital Assistant (in short PDA) because the mobile phone has been universally adopted for everyday use and because it offers both computational power and network connectivity. Additionally, STAMPS uses the cellular identifier provided by the wireless infrastructure to compute location.

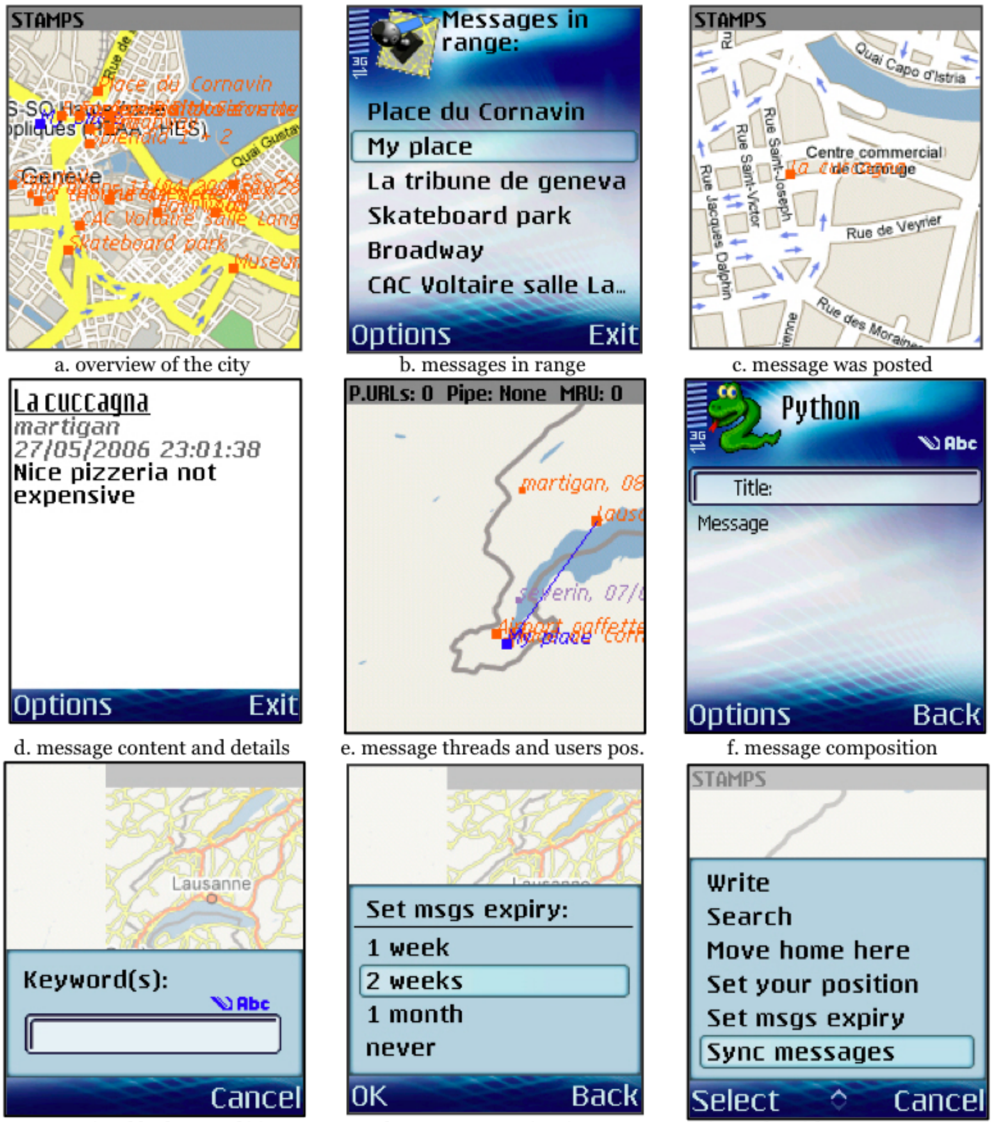
User interaction

STAMPS combines two main functions: it allows the user to visualize the maps of the place where she is located and to annotate these maps. Maps are streamed from GoogleMaps. Users can scroll the map, and zoom in and out of the available levels of detail (part (a) of figure 5-3). To annotate, the user can locate a specific point on a map and associate a message to it. Once posted, a square landmark is shown on the chosen position (part (c) of figure 5-3). To retrieve the content, the user can choose between two different strategies. The first strategy consists of zooming to a desired part of the city, and then displaying a list of the available messages (part (b) of figure 5-3). The second strategy consists of retrieving content through one or more keywords (part (g) of figure 5-3). This results in a list of messages as in the previous scenario, however these messages can refer to disparate zones of the map.

The message list displays only the message titles. The body of the message, as well as other details, are displayed only upon request. Part (d) of figure 5-3 shows the message content, the

[1] Several colleagues contributed to the development of STAMPS: Siddarth Jain, Shuja Parvez, Jürgen Scheible. See the acknowledgment section for more details.

[2] ShoutSpace was developed by several colleagues at EPFL: Fabien Girardin, Rachna Agarwal, Patrick Jermann, and Fabrice Hong. See acknowledgment section for more details.



La cuccagna
martigan
27/05/2006 23:01:38
Nice pizzeria not expensive

a. overview of the city

b. messages in range

c. message was posted

d. message content and details

e. message threads and users pos.

f. message composition

g. retrieval by keyword(s)

h. messages expire time

i. synch. with remote server

Figure 5-3: STAMPS display captures of major functionalities: (a) overview of Geneva with all the messages posted; (b) list of the messages visualized in the current map; (c) display of a message overlaid on the map: the anchor point is a small red square; (d) the content of a message is displayed; (e) replies to existing messages appear connected with lines; (f) a message posting encompasses a title, visualized on the map, and a body containing more detail; (g) messages could be retrieved by keyword search; (h) messages could be filtered by time to avoid display cluttering; finally, (i) synchronization with the remote server was a manual process

time and date at which it was posted and the username of the author. The author of a message is the only person entitled to delete a message from the system. More precisely, the message is flagged as deleted. It is not removed from the database but it is not synchronized any more with the other clients. Users can further decide to post a reply to an existing message. This results in a line visually joining the landmarks of the two messages (part (e) of figure 5-3). Additionally, the user can decide to disclose her position to a group of friends. This information is then shown on the map with the name of the user and the date on which it was provided (part (e) of figure 5-3).

Other commands allow the filtering of content based on a temporal criterion (part (h) of figure 5-3). This last option was designed to allow the user to avoid the cluttering of the display with message landmarks (see for instance part (a) of figure 5-3). The synchronization with the remote server is initiated manually by the user (part (i) of figure 5-3).

Finally, messages are coloured according to whether they belong to the user and whether or not they have been read. Messages belonging to the user are coloured in red, while messages of other participants are coloured in blue. Dark blue was used for unread messages, while light blue was used for messages that had been read by the user.

System architecture

STAMPS uses a client-server architecture. STAMPS stores user places and messages in a database on a server. This enables easy synchronization between the phones and the web-based clients. Additionally, to minimize data transfer, STAMPS keeps a local copy of the database in the phone's memory card. When asked to do so, the phone client pulls new messages posted in the system over a wireless HTTP connection and synchronizes the local database with a remote one. New messages are pushed to the remote database at message completion. At logout or at user discretion, the user's logs are sent to a remote repository via the same mechanism.

The software was implemented with Python for the Symbian platform[3]. It consisted of a single script that was executed through a shortcut icon listed in the main menu of the mobile phone. During the first execution, the software created a repository folder on the memory card of the phones where all the relevant data were stored: a) the map tiles streamed from GoogleMaps were cached locally to save connection airminutes; b) the application generated a local copy of the remote database where all synchronized messages were stored for offline browsing; and finally c) a single user log was generated at each login and copies kept on the memory card.

[3] This is an open source port of Python for Symbian series 60 phones and it is based on Python version 2.2.2, see `http://opensource.nokia.com/projects/pythonfors60/`, last retrieved April 2008.

Place acquisition

Users use the maps provided by the interface to select the location that is most meaningful for their current message. A pointer allows this selection. Once entered, the pixel coordinates on the phone's screen are converted to the corresponding latitude and longitude, associated to the entered text, and finally stored in the database. Conversely, for the actions produced in the system, I used the rough positioning information provided by the cellular network. Each time a user produced an action in the application, this was recorded in the logs (see section 6.1.3) and associated with the identifier of the GSM antenna to which the user was connected.

It must be noted that automated place acquisition is not the focus of this study. Other researchers are investigating this issue (Zhou et al., 2005). However, I think automatic place acquisition is one of the central issues in the development of location-based services, as it can reduce user effort during interaction with the system and has important implications in the cognitive processes resulting from the interaction with the system, as shown by Nova (2007).

5.3.2 ShoutSpace

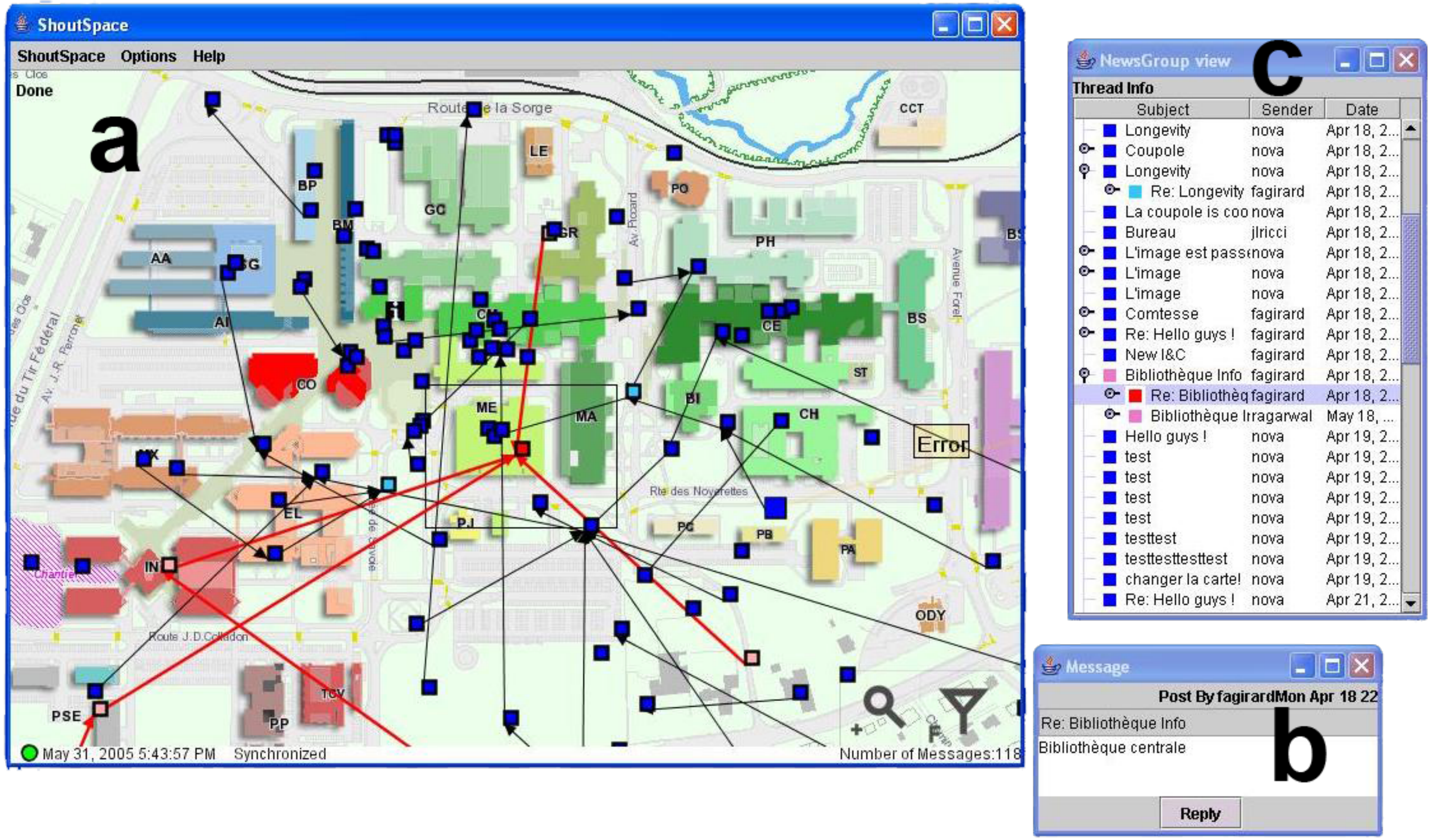
ShoutSpace enables users with WiFi mobile devices (PDAs and notebooks) to see their position on the EPFL campus and to leave messages to other users. The message is then dispayed on the map at a desired location. Multiple threads are displayed graphically with connections among the messages.

User interaction

The ShoutSpace interface is composed of three windows. The first window contains the map of the campus (part (a) of figure 5-4). The second window presents the message content (part (b) of figure 5-4) and the third lists all the messages available in the system ordered according to the tread to which they belong (part (c) of figure 5-4).

The map pane of ShoutSpace allows the user to navigate and access the messages left in the system. The core of the interface is the map of the EPFL campus. There users can see the messages left by other members represented as squares. The threads between the messages are represented as lines connecting the squares. Small arrows on the lines represent the 'parent-child' relation between them. Clicking with the mouse on an empty point of the map causes the map to be recentered at that point. The same effect can be obtained using the arrow keys. Moving the mouse over a square, will cause the message title to be displayed over the square. Clicking on a square will cause the message content to be visualised on the message window (part (b) of figure 5-4).

On the lower-right corner, two icons allow further interaction with the map: a magnifying



Thread Info

Subject	Sender	Date
Longevity	nova	Apr 18, 2...
Coupole	nova	Apr 18, 2...
Longevity	nova	Apr 18, 2...
Re: Longevity	fagirard	Apr 18, 2...
La coupole is coo	nova	Apr 18, 2...
Bureau	jlricci	Apr 18, 2...
L'image est pass	nova	Apr 18, 2...
L'image	nova	Apr 18, 2...
L'image	nova	Apr 18, 2...
Comtesse	fagirard	Apr 18, 2...
Re: Hello guys !	fagirard	Apr 18, 2...
New I&C	fagirard	Apr 18, 2...
Bibliothèque Info	fagirard	Apr 18, 2...
Re: Bibliothèq	fagirard	Apr 18, 2...
Bibliothèque	lragarwal	May 18, ...
Hello guys !	nova	Apr 19, 2...
test	nova	Apr 19, 2...
test	nova	Apr 19, 2...
test	nova	Apr 19, 2...
test	nova	Apr 19, 2...
testtest	nova	Apr 19, 2...
testtesttesttest	nova	Apr 19, 2...
changer la carte!	nova	Apr 19, 2...
Re: Hello guys !	nova	Apr 21, 2...

Post By fagirardMon Apr 18 22

Re: Bibliothèque Info

Bibliothèque centrale

Figure 5-4: Interface of ShoutSpace: (a) ShoutSpace Map window; (b) message window; (c) newsgroup view window (this pane was not used during the experiment reported in chapter 7)

glass and a funnel icon. The magnifying glass allows the user to move to a higher zoom level of map resolution (there are only two levels). Clicking the icon again will restore the visualization to the lower level of detail. The same effect can be achieved using the '+' and '-' keys on the keyboard. The funnel icon allows the user to the filter the available message. Its interaction mechanism is explained below.

The message window allows the user to read messages and to write replies to subjects of interest. On the main line the username of the author is reported and the time when the message was posted in the system. On the second line the title of the message is reported. In the body of the window, the user can see the actual content of the message. Clicking on the 'Reply' button of this window allows the user to select a new location where the content of the reply can be attached. The newsgroup view window orders the messages posted in the system according to the discussion threads to which they are associated. The messages in this window are sorted by date, with the most recent on top. However, this thread pane was not displayed during the controlled experiment reported in chapter 7.

The user can add new content to the system as a new message, which will constitute a new thread, or as a reply to an existing message. In the first case, she has to select the anchor point of the message on the map and then right-click and select 'new message' from the contextual menu. Subsequently, an empty message window allows the user to enter the title of the message and its content. To reply to an existing message, the user needs to select the original message, and click on the Reply button on the message window.

In ShoutSpace the words 'filter' and 'search' have equivalent meanings. It is possible to highlight messages on the map, matching messages by keyword or by time. To activate a filter the user has to select the funnel icon, or the corresponding 'F' key on the keyboard. Once a filter has been activated, the matching message(s) results are active while the others are fade out and inactive on the background.

System architecture

ShoutSpace mainly targets semi-mobile (e.g Notebooks) and mobile (PDAs) devices for the EPFL community. It is based on a client-server architecture. The server runs on Jakarta Tomcat. Clients synchronize their positions and the messages at regular interval. The notebook version, ShoutSpace, is written in Java and the PocketPC was programmed in .NET and C++. SOAP was a natural choice for a communication layer to allow proper interoperability. The automatic positioning algorithm uses the measures collected by Place Lab native libraries[4]. The nearby WiFi beacons' MAC addresses are matched to a 2D position and a centroid algorithm is used to calcu-

[4] Place Lab is an open source framework for device positioning. See http://www.placelab.org/, last retrieved April 2008.

late the approximate position. The semi-mobile version runs on WindowsXP, Linux and MacOSX, and supports many wireless network adapters.

5.4 Summary

This chapter presented the research questions that drive this thesis work. The chapter also explained the research methodology adopted, which consists of three experiments conducted in two different contexts and employing two custom-made CAS tools. The first experiment will be conducted in the field using a qualitative approach consisting of deploying the prototype application and recording, and subsequently analysing, the users' interaction in the system. This initial research will allow me to investigate the relations of annotations, the actual locations to which these refer and the map that will be used as a support during their interaction. The second experiment will research more specific hypothesis that were suggested by the review of the relevant literature. Testing these hypotheses required a quantitative approach and experimental design.

The research framework described in chapter 2 will be used to describe results from three experiments focusing on the influences of Collaborative Annotation Systems that implement Explicit Referencing mechanisms on collaborative tasks. The first two experiments will present qualitative observations on the use of a custom made ubiquitous CAS tool in a city-wide environment. These results helped to frame the scope of the third experiment, which was conducted in laboratory. In this last case, the literature review presented in the previous chapters led to posit that the availability of ER could have a positive effect on collaborative work at a distance. However, a second relevant dimension was identified in the messages organization criterium. Table 5.1, at page 143, summarizes the research questions, the related studies that will be addressed the questions and the associated hypotheses.

Chapter 6

Qualitative Observations of Map Annotation in the City

This chapter describes a field experiment that I conducted to answer the first three research questions. It reports observations from a trial where participants could produce location-based annotations using a mobile application and without a specific scenario of use. Additionally, the same system was used by urban planners, in the second study, to answer the fourth research question.

6.1 A twofold study using STAMPS

Chapter 4 reviewed a number of projects on connecting information to geographical positions, such as GeoNotes (Persson & Fagerberg, 2002), ActiveCampus (Griswold et al., 2003), E-graffiti (Burrell & Gay, 2002), and UrbanTapestries (Lane et al., 2005). All these prototypes allow users to express opinions, preferences, recommendations, questions, and jokes, connected to specific places. More recently, commercial companies have launched similar services that received extensive media coverage (Morgan, 2005; Meyer, 2004). Services like DisMoiOu[1] or Socialight[2] offer this kind of map plus content mash-ups. However, little research has been done on the user experience of these systems in the real-world urban environment. Many questions regarding the production and the consumption of information in such systems are still open.

The scenario under which many ubiquitous CAS have been built describes an user walking in the city and discovering a local resource that she did not know before. Then she decides to share this information with her group of peers. Therefore she creates a location-based message

[1] See http://dismoiou.fr/, last retrieved January 2008.

[2] See http://socialight.com, last retrieved January 2008.

pointing to this resource and posts this information to the CAS system.[3] Behind this simple scenario there are a number of assumptions: first, that the user is actually moving in the city space near the described resource at the moment in which the note is created; second, that she did not know of this resource before and it is discovered at about the time the message is posted. Finally, another hidden assumption of this scenario is that the annotation tool provides support to an activity for which there are no alternative practices in place. However, this might not be the case for the spontaneous use that users might do of such a technology.

These technologies are somewhat immersive: the user is located in the physical context that she describes, when she describes it. This is certainly an added value, but in many circumstances, authors may also annotate a place from a map, without physically being there. This may especially be the case if they know very well this place.

6.1.1 Rationale of the studies

The aim of this chapter is to better understand how annotations tools can be used to support human communication in a city wide environment. I investigate *how* and *why* users of an ubiquitous Collaborative Annotations System could produce and share geographically-enriched messages. Precisely, I address here the following research questions:

Q1 What kind of messages do users associate to map locations?

Q2 What is the relationship between the messages and the actual locations to which these messages refer?

Q3 How does the map mediate this relationship?

Answering these questions is difficult because research is conducted in an absence of established social practices: these location-based applications support new types of mobile experiences that are not yet common in people's lives. Therefore, as suggested by Crabtree (2004), I ran this field trial as a *breaching experiment*, observing whether and how this new technology could provoke practices and reveal contingencies between activities and technological interactions. In particular, previous ubiquitous CAS tools lacked a detailed logging report that could help to observe more systematically trends in the way participants used the application and that could prevent some of the *self-reporting bias* typical of interviews or questionnaires[4]. There were some exceptions. Systems like ActiveCampus or E-Graffiti supported detailed logs. However, these projects were developed in small geographical environments, namely campuses, with homogeneous users

[3] This scenario can be considered as the conceptual framework under which the qualitative observations reported in this chapter were conducted (Miles & Huberman, 1994).

[4] I am not arguing that traditional qualitative techniques are not valuable, and in fact I am employing interviews and questionnaires in this very chapter. Here, I am combining these traditional qualitative methods with the analysis of user-generated logs.

groups. The experiments reported in this chapter seek to overcome some of the limitations of previous research, by deploying a CAS system on a large city scale and with a detailed and integrated logging system.

6.1.2 Apparatus

Each participant received a mobile phone (maker: Nokia, model: 3320) with the installed STAMPS application. As I did not want them to use a distinct mobile for their personal communications, I encouraged them to put their SIM card inside the provided phone and use only that device for the period of the trial. This maximized the chances of having the application readily available when the need for it arose. Additionally, participants received a pocket user-guide with the instructions on how to operate the applications (see figure A-4, at page 330). The phone manual was available for download on the project website.

The reason why I had to provide a mobile phone to each participant is related to the specific implementation of STAMPS, which was, in fact, running only on Symbian-compatible[5] phones. This platform covers only 5% of the population in Europe. The selected participants did not have a compatible phone.

Figure 6-1 shows the phones that were used during the field trial. The codes on the back were added for the second experiment (described in section 6.3, at page 173), as participants were supposed to use this phone to coordinate their professional activities and at the same time they had their personal mobile phone for their private communications. Therefore for the second field trial, I assigned each participant a secondary mobile number. As remembering a new mobile number is sometime tedious, these numbers were stored in the address book of the phone of each participant associated with the color code marked in the back of the phones. The color codes could be an easier way to remember the mobile numbers of the other participants. Their aim was to help coordinate the exchange of pictures and other information when face-to-face or at a distance.

6.1.3 Measures

The STAMPS application generated fine-grained logs of the user's interaction with the system. Each action in the application was recorded and associated with a timestamp and extra detail customized to every kind of action. For instance, a 'zoom' command in the application was also associated with the coordinates of the resulting portion of the map that was displayed after the command execution. A 'read' command would have been associated with the unique identifier of the message retrieved. At the end of each session in the system, the logs were sent from the

[5] See `http://www.symbian.org/`, or `http://www.s60.com` for a list of compatible phones. Last retrieved June 2007.

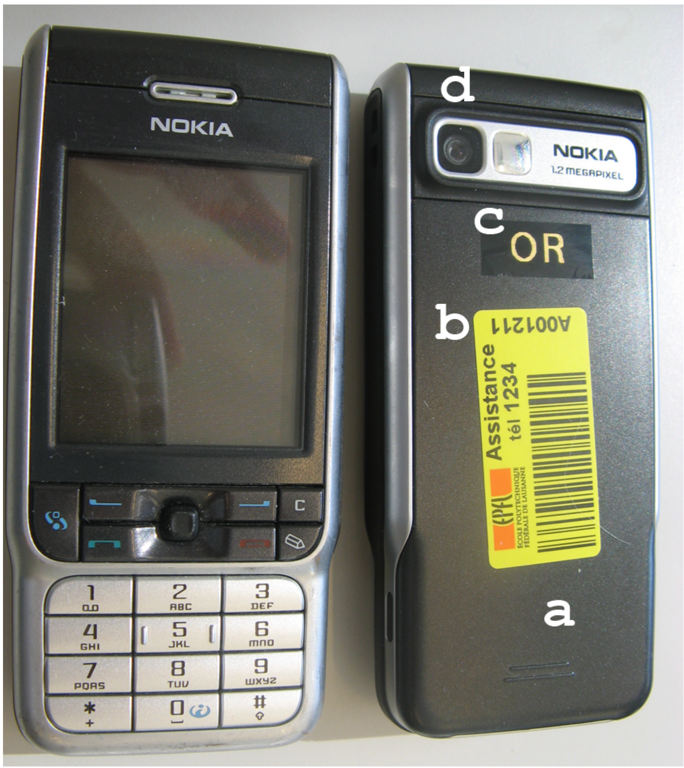
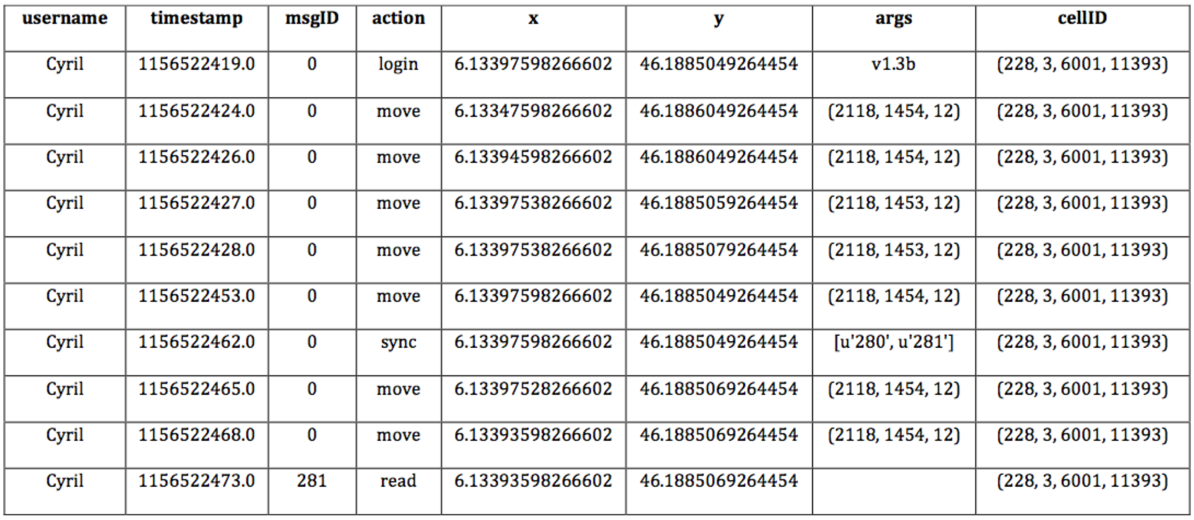


Figure 6-1: Example of phone used during the field trial. (a) phone serial number; (b) Bluetooth identifier; (c) color code added for the second field trial matching the mobile number of the phone in the address book ('OR' in French 'gold'); (d) camera lens

Table 6.1: Excerpt from the log files of STAMPS



username	timestamp	msgID	action	x	y	args	cellID
Cyril	1156522419.0	0	login	6.13397598266602	46.1885049264454	v1.3b	(228, 3, 6001, 11393)
Cyril	1156522424.0	0	move	6.13347598266602	46.1886049264454	(2118, 1454, 12)	(228, 3, 6001, 11393)
Cyril	1156522426.0	0	move	6.13394598266602	46.1886049264454	(2118, 1454, 12)	(228, 3, 6001, 11393)
Cyril	1156522427.0	0	move	6.13397538266602	46.1885059264454	(2118, 1453, 12)	(228, 3, 6001, 11393)
Cyril	1156522428.0	0	move	6.13397538266602	46.1885079264454	(2118, 1453, 12)	(228, 3, 6001, 11393)
Cyril	1156522453.0	0	move	6.13397598266602	46.1885049264454	(2118, 1454, 12)	(228, 3, 6001, 11393)
Cyril	1156522462.0	0	sync	6.13397598266602	46.1885049264454	[u'280', u'281']	(228, 3, 6001, 11393)
Cyril	1156522465.0	0	move	6.13397528266602	46.1885069264454	(2118, 1454, 12)	(228, 3, 6001, 11393)
Cyril	1156522468.0	0	move	6.13393598266602	46.1885069264454	(2118, 1454, 12)	(228, 3, 6001, 11393)
Cyril	1156522473.0	281	read	6.13393598266602	46.1885069264454		(228, 3, 6001, 11393)

phone to the remote server. Table 6.1 presents an excerpt of the log where the user 'Cyril' first log into the system, then moves five times over the map. Next he synchronize the logs with the remote server, recenters the map around a specific message and then read this message. A longer excerpt is reported in appendix B, at page 331, where all action categories are listed.

6.2 Study 1

During the summer of 2006, I organized a field trial of geographical messaging in Geneva, Switzerland. Participants were asked to use STAMPS, an application for mobile phones (see section 5.3.1). The Geneva location was chosen because of the relative proximity to EPFL (for logistic reasons) and because, at the time of the trial, it was one of the few Swiss areas covered by high-resolution imagery by GoogleMaps[6], the cartographical service that I used to provide the maps annotated by the participants.

6.2.1 Participants

I recruited participants through leaflets posted in shopping malls and universities (see figure A-3, at page 329). I also posted a call for participation on our university blog. Twenty-one people, 5 women and 16 men, from three different work contexts, volunteered to participate in the experiment, with ages ranging from mid-20s to the 50s. Fifteen volunteers were students from two distinct academic research groups in Geneva. Four other participants were journalists from a local newspaper. The last two participants worked for a pharmaceutical company in Geneva. They were all native French speakers. Also, they were all living in the center of Geneva for at least a couple of years, therefore they all knew the city well. The center of Geneva contains a mix of single and multiple family housing, restaurants, shops and other businesses, as well as schools and community centers. Many residents live within walking distance of retail districts. Area residents commonly use public transport and bicycle for negotiating the city center. None of the participants knew members of the other groups participating in the field trial but they all were familiar with their participating colleagues.

Each subject was reimbursed the connection costs that she had to sustain to connect to the remote server during the period of the trial. Also, as an incentive I informed each group that the most active participant would receive a prize of 100 Swiss Francs ($\sim$ 69 Euros).

6.2.2 Procedure

The phones were delivered during the first days of June. To get started, I met with each participant for a briefing. During this session, I explained how to operate the phone and the application. Each

[6] See `http://maps.google.com`, last retrieved February 2008.

participant received full support for installing her SIM card in the mobile phone, and explanations on how to copy contacts' details. These sessions lasted 45-60 minutes. As I wanted to leave the user free to use the system in any way she considered useful, no structured instructions of what to do with STAMPS were given. Instead, they were told to use STAMPS just as they wanted. To this end, the only information they received was the list of generic scenarios of use that was included in their pocket manual. Also the application was bootstrapped by posting an initial batch of 80 messages in the database containing information about entertainment places in the city center: restaurants, cinemas, hotels, etc. This information was extracted from a local tourist guide.

During the period of the trial, a help desk was available to troubleshoot problems, however this service was not used by participants. Additionally, participants had at their disposal a web site[7], where they could find a Frequently Asked Question guide, and a forum to discuss possible issues. Using the web sites, participants could download the messages posted by their group to their computer in a format compatible with mapping applications like Google Earth[8]. Additionally, during the three month trial period, regular updates were sent on the status of the project and checked that the participants were still interested in continuing. Occasionally, these messages included statistics of messages posted in the system.

At the end of the study, I conducted in-depth interviews with each subject about their experience with STAMPS and I administered a questionnaire to those who could not meet face-to-face. This exit interview lasted about an hour and concerned the way in which participants used the system. I was particularly interested in understanding whether they found useful information in the system and how they used this information. Finally, I tried to ask side questions on their privacy concerns and their willingness to publicly share their messages.

6.2.3 Results

Overall, users created 162 map annotations (see the annotated area in figure 6-2), including new messages and replies. Twenty-one participants used the system during three months. The application generated a log file each time an user logged in the system, independently of whether or not they posted a message. Therefore, I could collect 866 log files accounting for about 734 hours of logging time ($\sim$ 30 days). The specific design of the application allowed me to record the users' interactions with the system to a high level of detail. I analyzed the collected messages, the locations where they were produced, and the logs generated during the interactions. In addition, I interviewed and administered a post-trial questionnaire to the most active participants.

I categorized participants into three groups based on their usage patterns: (P) *the passives* (5

[7] The project web site is available at `http://www.shoutspace.eu/`, last retrieved August 2007.

[8] See `http://earth.google.com/`, last retrieved August 2007.

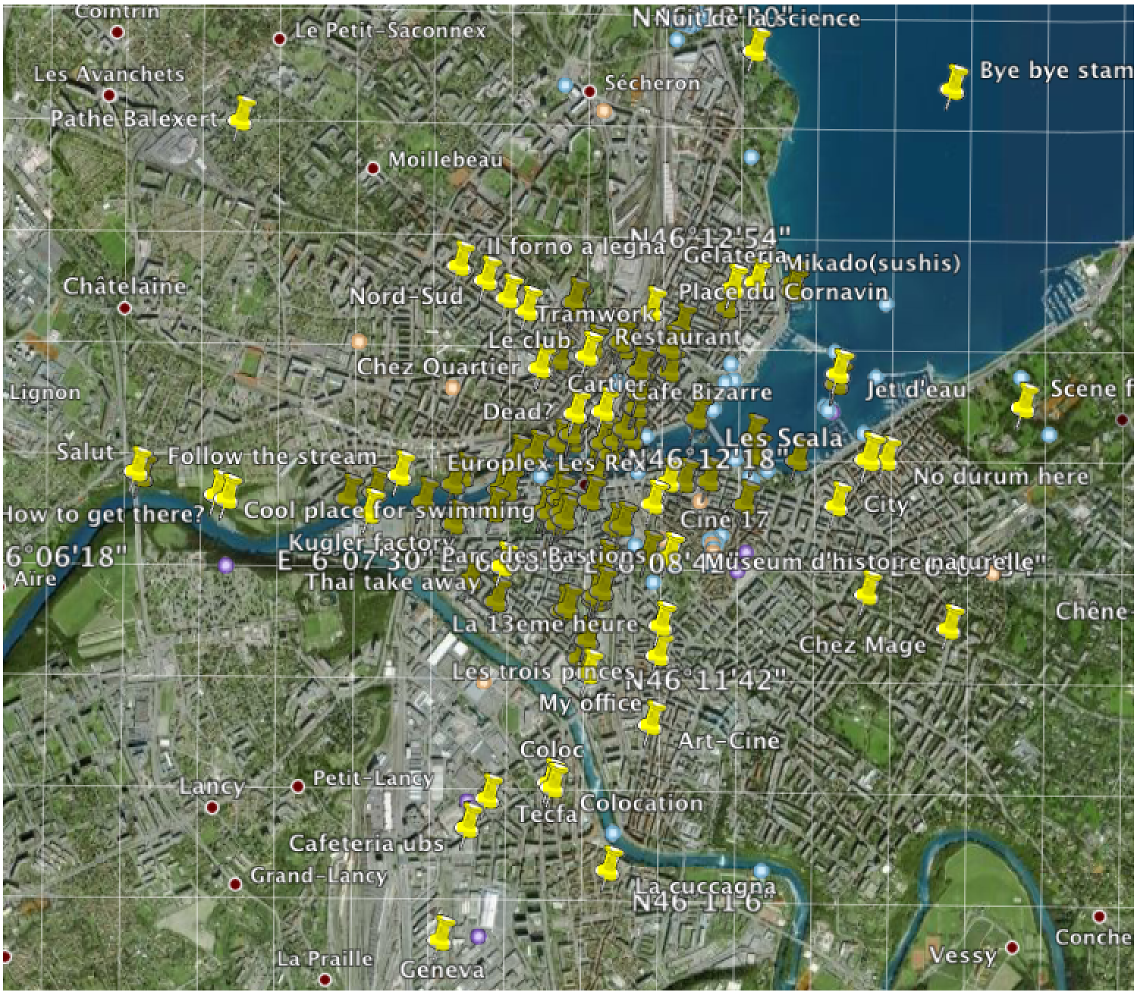
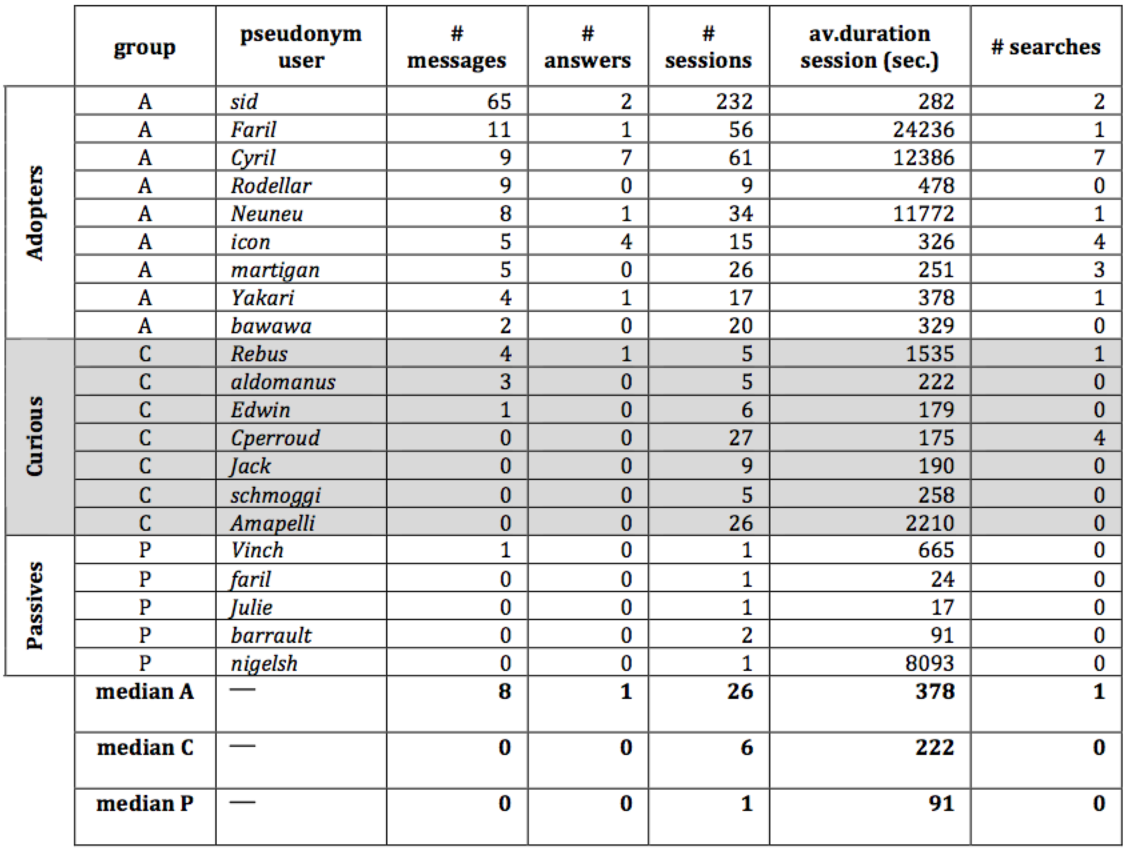


Figure 6-2: Annotation area of the first field trial. Pinpoints represent messages (© Google Maps image, Google Inc.)

participants), those who logged in the system only once and did not produce any contribution; (C) *the curious* (7 participants), those who participated briefly in the activities posting one or two messages and logging in an average of 5 times; and finally (A) *the adopters* (9 participants), those who logged into the system frequently, often leaving their application running for a long time, produced most of the messages and engaged in many conversations. Table 6.2 shows some quantitative data of the analyzed dataset. The group of adopters included users who were particularly enthusiastic about mobile technologies. This is a conjecture that I drew from the fact that these participants all had a blog and commented regularly on new technologies. Other than this fact, I could not find any differences among participants who fell into the curious group and the adopters, nor any commonalities among 'curious' participants given the demographical information that I could obtain from the interviews and interactions with them.

The group of passives was composed of three journalists and the two participants from the pharmaceutical company. This unenthusiastic group of users was nearly 25% of the study population. It could have been interesting to understand the reasons behind the lack of participation of the passives, however it was extremely difficult to obtain more information from these participants. They claimed to be extremely busy and after an initial demonstration of interest for the experiment, they declined further invitations to interviews and questionnaires.

Table 6.2: General statistics for three months of system usage



	group	pseudonym user	# messages	# answers	# sessions	av.duration session (sec.)	# searches
Adopters	A	*sid*	65	2	232	282	2
	A	*Faril*	11	1	56	24236	1
	A	*Cyril*	9	7	61	12386	7
	A	*Rodellar*	9	0	9	478	0
	A	*Neuneu*	8	1	34	11772	1
	A	*icon*	5	4	15	326	4
	A	*martigan*	5	0	26	251	3
	A	*Yakari*	4	1	17	378	1
	A	*bawawa*	2	0	20	329	0
Curious	C	*Rebus*	4	1	5	1535	1
	C	*aldomanus*	3	0	5	222	0
	C	*Edwin*	1	0	6	179	0
	C	*Cperroud*	0	0	27	175	4
	C	*Jack*	0	0	9	190	0
	C	*schmoggi*	0	0	5	258	0
	C	*Amapelli*	0	0	26	2210	0
Passives	P	*Vinch*	1	0	1	665	0
	P	*faril*	0	0	1	24	0
	P	*Julie*	0	0	1	17	0
	P	*barrault*	0	0	2	91	0
	P	*nigelsh*	0	0	1	8093	0
	median A	—	**8**	**1**	**26**	**378**	**1**
	median C	—	**0**	**0**	**6**	**222**	**0**
	median P	—	**0**	**0**	**1**	**91**	**0**

Therefore, the analysis reported below was focused on the last two groups, looking for differences in their login sessions, their consumption and production style in the system. the logs were analyzed looking for patterns. The results described in this section are prototypical situations, or "vignettes" that could account for the most frequent situations (Kirk et al., 2005). Of course the proposed styles of use should be intended as qualitative descriptions and should not be understood as quantitative differences between the two groups. To analyze such complex data, I built a timeline visualization that shows the time interval and the sequence of actions that the users performed as derived from the log files. In these visualizations, circles represent zoom actions, while triangles are steps towards one cardinal direction. Squares represent a 'read' action (e.g., when a message is retrieved for reading). Rectangles are retrieval by keyword/s. In the last part of this section, the produced messages were categorized according to two coding schemes.

Login sessions

One of the questions that the logs helped to get answered is: when did participants use the application? And for how long? On average, the adopters logged into STAMPS 52 times (min. 9, median 26, max. 232 times) and their sessions lasted 93 minutes (min. 4, median 6, max. 404 minutes), while the curious logged in 11 times (min. 5, median 6, max. 27 times), and their sessions lasted 11 minutes (min. 3, median 4, max. 37 minutes). Looking at the individual login sessions, the 'curious' used the system mostly with a 'browsing' attitude, moving around rapidly through the tiles of the map, zooming in and out and rarely taking time to read messages and to perform searches in the database. On the contrary, 'adopters' allowed time between each action for the application to correctly load the map tiles. They choose the regions to explore with care with a few clicks and moves and finally posted messages and read available contributions. Figure 6-3 offers a comparison between these different attitudes.

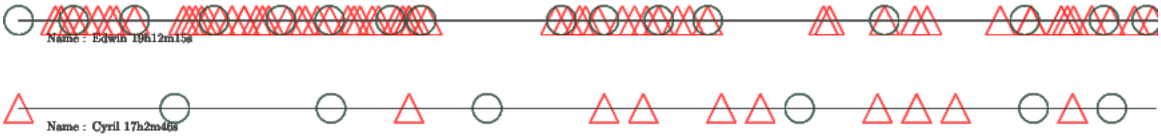


Figure 6-3: Timeline comparison of login sessions. An 'adopter' on the bottom and a 'curious' on the top. Circles represent zooming in or out the map, while triangles are moves in the four cardinal directions

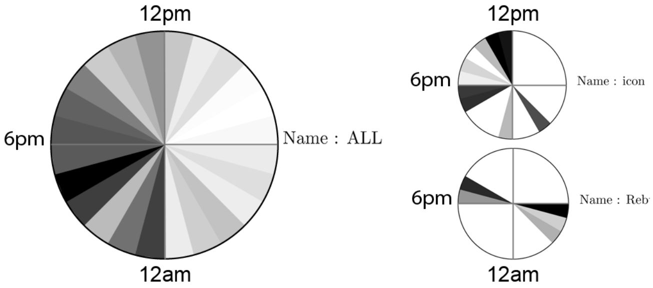


Figure 6-4: Left, cumulative representation of the login hours for all the participants. Right-top, login hours for an 'adopter'. Right-bottom, login hours for a 'curious'

Additionally, I looked at the hours of connection for each participant. I found that the whole population used STAMPS early in the morning, immediately during lunch break and after 5 pm. Figure 6-4 shows usage patterns for the 'adopters' and the 'curious' and a cumulative representation of the login hours for all the participants. It must be noted that this representation is not a standard pie chart. It represents a 24-hour clock. Each sector corresponds to one hour of the day. The color of the sector corresponds to the relative frequency of logins at that particular time of the day. This gradient goes from white for zero logins to black for the maximum number of logins at a particular time of the day.

Information retrieval

One of the interests of this observation was to understand how participants retrieved the messages stored in the system. On average, 'adopters' ran two queries in the database (min. 0, median 1, max. 7 queries) during the three months of the field trial, while the 'curious' did not run any queries (min. 0, median 0, max. 4).

Most of the time participants used the map to navigate towards a region of interest first and then they red all the messages available for that particular area. In a minority of situations, users retrieved content searching for messages matching specific keywords. A 'search' function of the interface allowed the user to execute keywords-based queries. Search results could be messages 'attached' to distant geographical locations. Figure 6-5 shows the logs for these two behaviors.

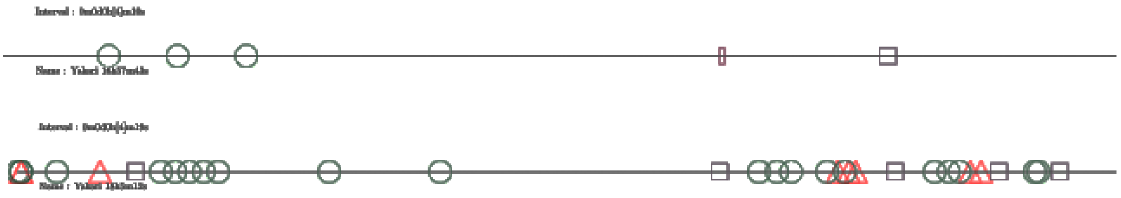


Figure 6-5: Timeline comparison of reading styles. Retrieval by content on the top, and retrieval by position on the bottom. Rectangles represent a search by keyword, while squares represent read actions

Production style

Was the participant standing at the location she wrote a note about? The logs helped analyzing the distance from message production to message content. At the exact time a message was posted, I logged the GSM[9] network cell identifier to which the mobile was connected. This is a unique number that distinguishes each cellular antenna worldwide. In a densely urbanized environment like Geneva, the radius of each cell ranges between 100 and 500 meters (∼ 109–547 yards). This information was used as a rough indication of the position of the emitter of a message. I found that all the messages in the database were posted from 49 different antennas.

[9] Stands for Global System for Mobile communications. It is the most popular standard for mobile phones in the world.

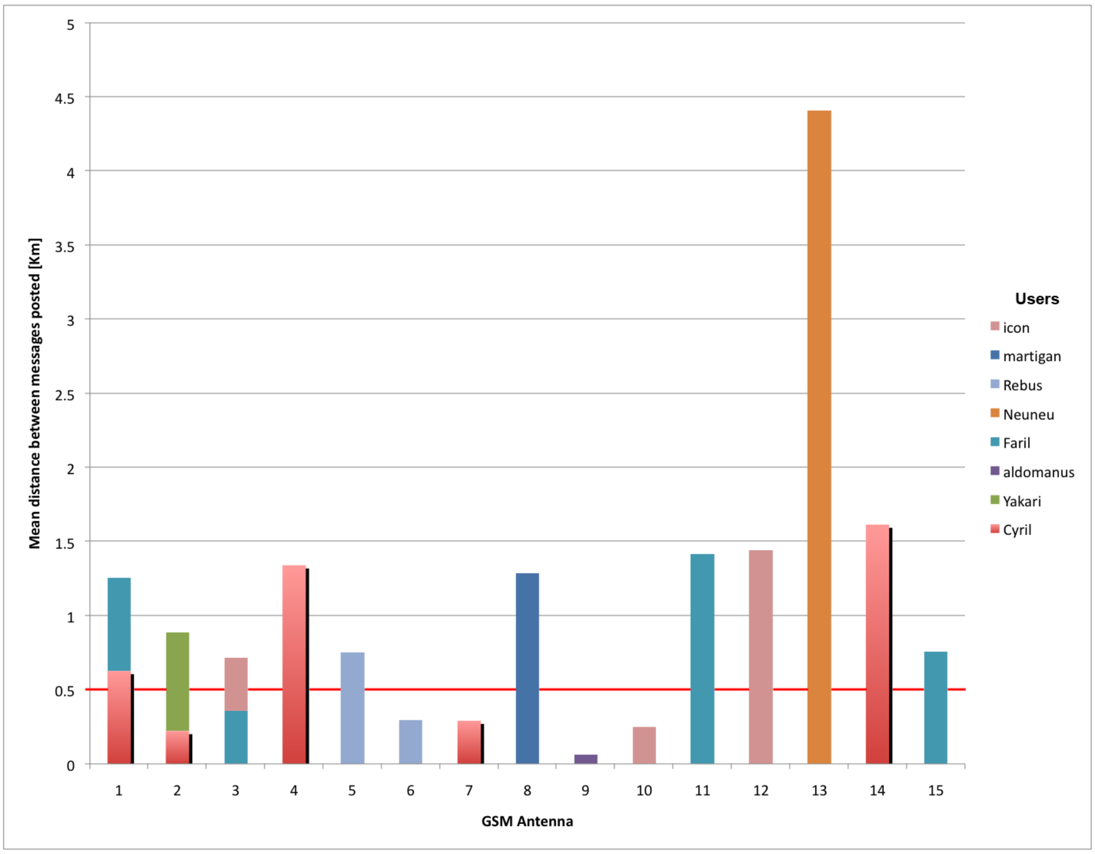


Figure 6-6: The height of each peak correspond to the average distance of the anchor points for the messages posted while users were connected to the same GSM antenna. The colors represent the proportion of message posted by each participants using that particular antenna. This graph only reports the antennas from which at least two messages were posted

A more detailed analysis revealed that participants used one to four different antennas to post their messages (see figure 6-6). This was a rough indication of position because the identifier of the antenna could not be mapped to coordinates in space (this information is kept confidential by the mobile operators).

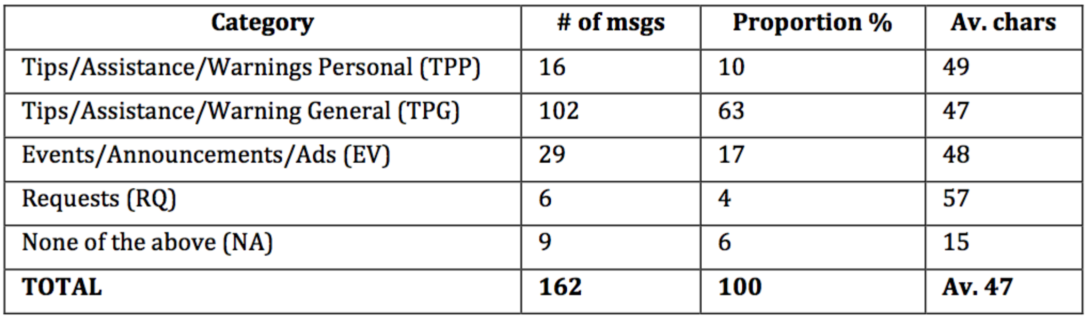
I therefore calculated the average distance of the messages posted using the same GSM antenna, under the assumption that if these messages were concerning events or items in the region covered by the antenna, then the distance of their anchor points should not exceed two times the radius of a GSM cell, namely 200 meters in a densely urbanized area. Since in most cases the distance between anchor points of message posted from the same antenna exceeded 500 meters of average distance, I concluded that users had a general attitude to publish content 'attached' to locations far from their actual position (see figure 6-6).

Content of the messages

In this section, I describe examples of messages produced during the trial. The messages have been translated as they were originally written in French. I categorized the messages posted during the trial using a coding scheme aimed at distinguishing the content of the contributions. As there is no widely accepted coding scheme for geo-localized messages, I defined my own, based on a discussion with other researchers building similar types of spatial annotation software (Tester, 2004). I used the five categories that I describe next. The main category is tips/assistance/warnings notes offering useful information for the reader at a particular location (e.g., "*Beautiful view of the lake from the bridge*"). I subdivided this category into personal notes (**TPP**) and general use notes (**TPG**), depending on whether the message was intended for a group of friends (e.g., "*This is the place where I work with Rork*") or to all the users of the system (e.g., "*Best pizzeria of the city*"). Messages in this category did not have a particular temporal validity. On the other hand, messages in the events category (**EV**) lost their value after a certain temporal window (e.g., "*In the afternoon the tram went off the rails. People are still working on it*"). I classed in this group advertisements for items on sale (e.g., "*I am selling my bike, 150 CHF*"); invitations to parties (e.g., "*The faculty fiesta is there tonight. Hope I'll have fun*"); concerts (e.g., "*Concert Saturday, Sand over skara, make it pink and honey for petzi. 21h Piment rouge, 10 CHF*"); and other entertainment events (e.g., "*Improvisation matches, from the 2 to the 11 of November 2006 www.impro.ch*").

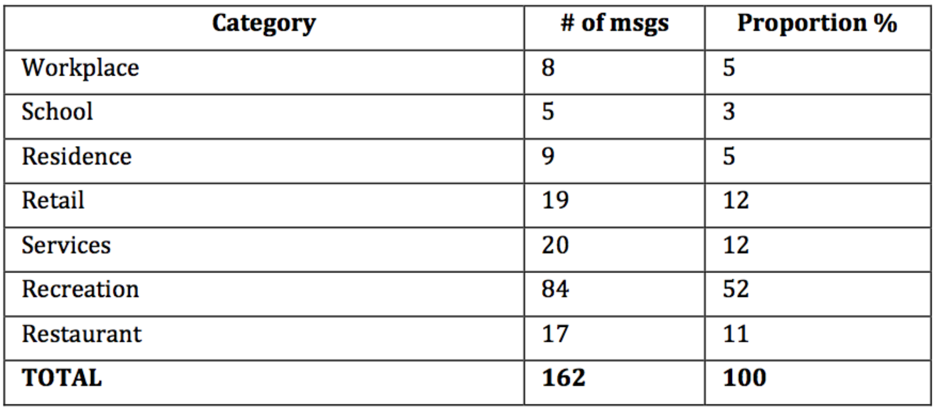
A different category describes 'spatialized' requests (**RQ**): people looking for a particular good or service in a specific spot of the city (e.g., "*A friend is looking for a roommate for 6 months. 700 francs/months. Call me!*"). Finally, I used a separate category for tests and messages that could not be coded with the above categories (**NA**) (e.g., "*Nice to meet you*"). Table 6.3 summarizes the frequency of each category of messages posted in the system during the trial. The last column of the table reports the average number of characters of the messages in each of the five categories.

Table 6.3: Number of messages produced for each category



Category	# of msgs	Proportion %	Av. chars
Tips/Assistance/Warnings Personal (TPP)	16	10	49
Tips/Assistance/Warning General (TPG)	102	63	47
Events/Announcements/Ads (EV)	29	17	48
Requests (RQ)	6	4	57
None of the above (NA)	9	6	15
TOTAL	**162**	**100**	**Av. 47**

Table 6.4: Categorization of places / messages (coding scheme inspired by Ludford et al. 2007)



Category	# of msgs	Proportion %
Workplace	8	5
School	5	3
Residence	9	5
Retail	19	12
Services	20	12
Recreation	84	52
Restaurant	17	11
TOTAL	**162**	**100**

To better understand the relation of the messages and the locations to which these messages refer, I categorized the messages using a second coding scheme that describes the places to which they are attached. I began with a similar classification developed by Ludford and collaborators (2007). They used a scheme containing 17 categories that was developed to classify messages produced with a location-based reminder application. However, I reduced the number of categories, grouping together all the retail types of locations and the service type of locations. I ended up with seven categories:

1. **Workplace** (e.g., "*My office: number 5158*", Vinch)

2. **School** (e.g., "*Uni-mail: Main university building*", Cyril)

3. **Residence** (e.g., "*My place: This is my room!*", Icon)

4. **Retail** (pharmacy, grocery, health care, hardware store, gas station, etc. E.g., "*Cartier: They have great chocolate rolls, also on sunday.*", Cyril)

5. **Service** (public service, train station, financial service, bank, church, airport, etc. E.g., "*Bus 10 Stop train station*", Rodellar)

6. **Recreation** (bar, plaza, beach, cinema, museum, sightseeing spot, sport facility, theater, stadium, etc. E.g. "*Parc des Bastions: A nice park where you can play big chess games*", Yakari)

7. **Restaurant** or pizzeria (e.g., "*Pizzeria Borgia: Plz avoid this pizzeria! Have gone there twice, two disappointments :(*", Neneu)

Table 6.4 reports the result of the coding of the messages following the presented schema. Results show that messages produced in STAMPS touched a variety of everyday places. Of course, I could have divided further the 'recreation' category to have a more homogeneous distribution however the main point of this result is that as STAMPS was intended mainly for public communication, people were reluctant to post information concerning their private life.

Interview and questionnaire analysis

At the end of the field trial, I scheduled interviews with three participants who were willing to chat about their experience and that had some free time. My interest was to complement the objective analysis performed on the user's log with subjective opinions on the usefulness of the application. Additionally, I sent a questionnaire to the others that could not meet in person. Six 'adopter' participants completed the questionnaire. One of the interviewed participant belonged to the 'adopters' group, while the other two belonged to the 'curious' group. The questions used in the interview were identical to those sent out with the questionnaire. In reporting the results, a neutral reporting was adopted.

In which situations do you think using messages that refer to specific locations in space, like those of STAMPS, might be useful? All the respondents answered that geo-located messages can be useful in situations where at least two persons want to communicate content that is related to a physical location. The situations most frequently suggested were: showing local recommendations or the history of a place; revealing personal footprints to friends; information about happenings like concerts, highway traffic, spaces left in a certain parking lot; ubiquitous games. One participant highlighted the fact that in order for the system to be useful its use should be related to an existing scenario, like a school trip or a work assignment. He specifically said that it might be interesting for learners to have information in context, which may lead to enriching discussions about buildings/monuments, and other elements in the physical landscape.

Interestingly, none of the participant interviewed looked at STAMPS as a system that could help find local resources. Participants explicitly said that when they are in the need of a local resources they use other solutions: (1) they look it up on the directories or else (2) they ask a friend. Alternatively, (3) some used social navigation (e.g., a line of people standing in front of a restaurant) or (4) to walked randomly through the retail area of the city. I also asked *how do they use services like GoogleMaps*. Some answered that these services are used for navigation but only when they do not know the place (e.g., going to a new city and looking for a specific building). I attempted to understand *what were the benefits or the limits of showing information on a map*. Many interviewees answered that reading a map is not obvious. Some had problems with maps, especially they complained of their difficulty in finding reference points for orientation. Conversely, other participants said that it is sometimes easier to navigate space with a map instead of a list of textual directions.

None of the participants mentioned the idea of location-based reminders that I described in chapter 4. When asked explicitly, some said that they did not think about it, but even if they did the application would have not support this function properly as it was lacking an alerting function for incoming messages.

How does STAMPS compare to other messaging systems like SMS or Newsgroups? Why/When would you use a SMS, an email or a news post instead of posting in STAMPS?
All the respondents noted that an SMS or messages on a forum are independent of location. They are one-to-one or one-to-many asynchronous conversations like messages on STAMPS but they do not relate specifically to space. Respondents described STAMPS as a communication tool in between SMS and newsgroup messages in terms of time-resources necessary to produce a message. While a news post is often long and detailed as directed to a wide community, an SMS can be quite informal, easy to compose and directed to a single person. STAMPS requires an extra effort to anchor messages to specific points on the map.

Do you remember any instance when you found something interesting/useful in the messages collected

in STAMPS?
One participant, originally from Bern, said that she found interesting tourist advices of Geneva. She found indications on the locations of the beaches on the lakeside, public baths, and nice sightseeing spots. All the other respondents answered that they could not remember any particular instance where the application was useful. They all wished the application could have contained advices in particular situations like a traffic jam or a strike, but it was not the case.

Have you ever written an article for Wikipedia or similar community-driven sites? Could you tell the story? Even if you did not, how would it compare to posting message in STAMPS? Most of the respondents did not have experience in writing contributions on a wiki. Two participants mentioned differences in the user interface: STAMPS does not have a reviewing process like Wikipedia. Also, it is easier to produce written contributions with a computer keyboard than with a mobile keypad. Finally, a participant mentioned that the lack of automatic positioning introduced mistakes in the actual location of the places to which the messages referred.

One of my interests was to understand *why people want to create location-based annotations*. Participants drew similarities of STAMPS with other social platforms like Flickr or Facebook, where people share content in order to define themselves. Somehow in these services, the benefit is given to the social status that you can obtain by sharing something particular, amusing or useful. Similarly, STAMPS might serve this function, but only if it is accessible to the group of peers (this is my interpretation of the various answers that I collected). In order for STAMPS to be useful, and used in social actions like petitions, it must reach a critical mass of users. Participants commented that in this regard, city activists are a minority and therefore they alone cannot propel the use of an application like STAMPS. This led me to another central question that I asked: *what is the value of local information?* Some participants answered that they look actively for regional information on local newspapers. They use this information to establish a sense of belonging with their group of peers, e.g., "to show that I share the same interests".

What is the biggest limitation of STAMPS? Why did you stop using it? In your opinion, why it did not work?
Three participants mentioned that they did not have enough location-based interests in common with the community of users that were testing STAMPS. Another point that was mentioned was the lack of richness in the database: participants mostly already knew the information that was posted in the system. One of the participants suggested a multi-modal solution to the problematic typing on the mobile phone. He suggested that messages should be produced on a desktop computer at home or at the office, while retrieval is fine with a mobile phone.

Several participants complained about a general lack of awareness about the life of their messages. They could not know whether a message was read and by whom, but also when they wrote something targeting a specific user, they could not know whether or not that user retrieved the message.

Did you had any privacy concern while using STAMPS? All the respondents agreed that they did not have such concerns as they knew that messages were visible by everyone and therefore they posted neutral messages to avoid any disclosure. In fact, many respondent said that they would have been willing to re-publish the messages that they posted during the trial on the internet or elsewhere as they did not contain any private detail.

6.2.4 Discussion

Q1, Which kind of messages do users associate to map locations?
I did not offer a structured scenario to follow and participants produced, most of the time, messages aimed at the whole population and not to a specific user. These findings are contrary to that of Persson and Fagergerg (2002), who found that notes were targeted at other participants for social communication. STAMPS was perceived more as a person-to-community asynchronous information tool than a person-to-person communication application. This result goes against Burrell and Gay's report on the use of the E-graffiti platform (Burrell & Gay, 2002). They found mainly a synchronous use of the messaging system. However, our system was missing a notification service for new messages. Thus, emitters could not have the certitude that a message sent to a specific person was read. I derived that, in order for STAMPS to function as a chat, recipients need to be notified promptly of incoming messages, and emitters need to be notified when their message was read. Therefore I can answer to the first research question by stating that STAMPS users, who were not given a precise task to accomplish, produced asynchronous messages aimed at a general public, and concerning recreation areas. However, the analysis also suggested that this spontaneous use of the tool might have been influenced by two specific characteristics of the system: first, the lack of an alert for incoming messages encouraged asynchronous use; second, as privacy was not granted by the system, users produced neutral rather than personal messages.

The analysis reported above shows that sharing location-based annotations with a mobile device is an emerging practice for which there are no established social conventions. I derived this conclusion from the answers given to my first survey question, in which participants listed a wide number of situations in which STAMPS might yield useful results. Many of these, like the social navigation of the city, are widely observable social practices (e.g., the number of cars parked in front of a restaurant, as well as the waiting line before a theatre indicate the places' popularity). However, I found little evidence of these practices performed in STAMPS. It is difficult to give a clear explanation of why participants suggested possible use of the tool that they did not employ. Nevertheless, questionnaire respondents explicitly stated that one of the reasons the application was uninteresting was because it was not available to their entire social network, or somewhat similarly, that participants did not have enough location-based interests in common. Many participants, especially the 'passives', did not see the utility of the service because it lacked content and perhaps usability (this conclusion was derived from interviewees' comments).

Q2, What is the relationship between the messages, and the actual locations to which these messages refer?
Annotating maps has been used as a leisure activity, performed during breaks or during commuting time; I did not relate it to applications to any profesional task that users had to accomplish. Contrary to my expectations, participants produced most of their annotations while far away from the actual locations to which their messages referred, as originally hypothesized. This led me to think that the content posted in the system was somehow familiar to the authors of the annotations and not discovered at their physical location. These findings were contrary to that of Griswold et al. (2003), who found that local context was very important for the content of the messages.

Local features of a certain place can be reconstructed in the mind and their subsequent narration derived from this mental image. Certainly, the opposite is also possible (from the actual situation to constructing the mental image). Perhaps being in the place during composition allows a higher fidelity and effectiveness in communicating about space. Understanding the subtle differences of geographical annotations produced in place versus those produced at a distance is not possible here given the experimental design that was adopted. To further understanding in this area is perhaps necessary a more controlled experiment specifically designed to compare linguistic features of location-based annotations produced in the field with those produced at a fixed location.

Q3, How does the map mediate this relationship?
I observed an *compelling effect* of the map in the user interaction with the application. Participants were 'attracted' by the map and spent most of the time just browsing tiles instead of looking for content. I made this conclusion from the small number of content-driven queries to the system and from the small number of discussions engaged during the trial. The system did not contain many messages. At the beginning of the trial there were only the 80 annotations initially used to bootstrap the application, while towards the end of the trial there were over two-hundred annotations. Even if there were not many messages to be looked at, users could not know in principle the current status of the database (how many messages) and therefore many queries returning no-results could have been executed. This was not the case. STAMPS followed a *map-first interaction paradigm*, in the sense that the map was used to route the user toward the content and not vice versa. While this 'map-first' approach supports users just browsing for interesting things nearby well, the mechanism is burdensome for those looking for specific and new content without a particular location in mind. These findings provides clues to answer the third research

question (Q3): the availability of a map influenced the way participants looked at the content available regardless of their actual context and particular needs. The results of this initial study do not allow for further understanding of the nature of this mediation. One can hypothesize that several kinds of mediation can result from different designs of the way messages lead to the map or viceversa, and therefore dissimilar outcomes in these diverse systems can be observed. This is going to be the objective of the experiment reported in the next chapter.

Overall, I attributed the differences of the presented results from the other studies of location-based annotations on a campus setting (Persson & Fagerberg, 2002; Burrell & Gay, 2002; Griswold et al., 2003) to the different geographical scales employed. While a university campus can be identified with a specific social group, like that of the students who inhabit it, a city space is 'impersonal' as it is being used by a multitude of different groups. It is therefore natural that a majority of messages were addressed to a generic community. Even if participants knew that their messages were mainly visible to their peers, this did not help them to see STAMPS as a place for interpersonal communications.

As I felt that a more specific scenario given to the users could play an important role in the obtained results, I designed a second field trial with students from an urban planning course at EPFL.

6.3 Study 2: Urban planners using STAMPS

The challenges that urban planning have to face are changing dramatically over the years. Cities and citizens' needs are becoming increasingly complex, and therefore designing living space requires a more analytical approach than the artistic or functional criteria that dominated this discipline in the past. Already in the 60s, Lynch was advocating for a more situated approach for urban planners. In his vision, empirical methods, such as questionnaires and interviews, could have been conducted to collect local data that was, in his view, fundamental to designing for urban areas.

Of course, Lynch was not alone in his perspective of what his profession should become. In the 50s, situationists lead by Guy Debord in Paris coined the term *dérive* to express their will to re-appropriate urban space. Dérive (literally: "drift") is a technique of rapid passage through varied ambiances. It consists in walking through the city, by being drawn by the attractions of the urban environment or the encounters they find there. Emotions are one of the most important means of selecting the way through the city. They created the term *psychogeography* in order to express "*the study of the precise laws and specific effects of the geographical environment, consciously organized or not, on the emotions and behaviour of individuals.*" (Debord, 1955). Situationists then came out with *psychogeographical games* like navigating through Paris with a map of London.

Developing analytical skills was also one of the objective of the P*roject Urbain, Mobilité et Environment* (urban project, mobility and environment, in short UE-C), held in the winter semester 2007 at EPFL. The course had two objectives: on one side, it aimed at training students in choosing, collecting and analyzing quantitative data from different sources (e.g., demographics, employment, mobility networks, social surveys, etc.); on the other side, it aimed at developing students' skills in collecting, filtering and analyzing qualitative data, like architectural barriers visible on site, feelings or emotion in a particular place, or self-reported perception of noise levels or pollution.

The course proposed the study of the west of Lausanne[10], in Switzerland, a region that included seven villages with about 50'000 habitants. This area underwent rapid growth over the last years. The west of Lausanne takes advantage of a strategical position in the urban agglomeration of Lausanne, as it is close to the university district, and the highway network. The challenges targeted by the course were the development of specific strategies to diversify this sector, which hosts many industries and enterprises, but also to preserve and ameliorate the quality of life of its inhabitants.

As a part of their course, students walked repeatedly in this area of the city, either all together or in small groups. The objective of these walks was to collect qualitative data: subjective feelings or emotions that were related to the particular moment of the visit (e.g., light exposure of a children's park at 3pm) and local facts that were so fine-grained to be unavailable through institutional databases (e.g., misleading route signage). Students of previous years' courses used low-tech solutions to record these experiences. Either they relied on their memory, or they took notes using paper maps and pencils. Notes were taken with the objective of discussing the students' perceptions and reflections on the site in class. The objective of these annotations was also that of structuring a critical review of the status of the places and to elaborate an intervention to ameliorate the citizens' quality of life.

This seemed a perfect situation for testing the usefulness of STAMPS in a concrete and defined scenario. I framed my fourth research question as: *Q4, do the results of the first three research questions change when the participants use the CAS tool in a structured task?* My idea was that the mobile device could have been integrated to this learning activity to support the students' activities. In particular, it was my conviction that students could benefit from using STAMPS to take the notes on site and to share and coordinate this activity in real time with the rest of the class. Additionally, as the application allows for the export the produced annotations in different formats, it could have been useful to produce visualizations for the final report. Therefore to test this qualitative hypotheses, I introduced this tool during the UE-C course of the winter semester 2007.

[10] This area is comprehended between the coordinates `x:531000`, `y:151000` and `x:538000` and `y:158000`, defined according to the Swiss coordinates system CH1903.

6.3.1 Method

Students enrolled in the UE-C were assigned to different sub-projects concerning specific aspects of the area (e.g., mobility, shopping centers, etc.). There were nine groups in total but only four of them could use STAMPS as only 16 compatible phones were available at this stage for the trial. Therefore, sixteen students participated in the field trial. They were all about the same age (21-23 years), they all lived in Lausanne and they were all French-speaking. They belonged to different programs at EPFL: architecture (10 participants), urban studies (4 participants), environmental studies (1 participant), and geography (1 participant). They did not receive any monetary incentive for their participation, as the activity was somewhat related to their course requirements. They received a prepay SIM card, in addition to their own, with the starting credit of 20 Swiss Francs ($\sim$ 13 Euros), enough to cover the expenses of producing over 100 messages. They received the same phones as in the first trial. However a color-code was added in the back of the phone (part (c) of figure 6-1) to help the students quickly find each others' phone number in the address book. This time the system was not bootstrapped with an initial set of localized messages.

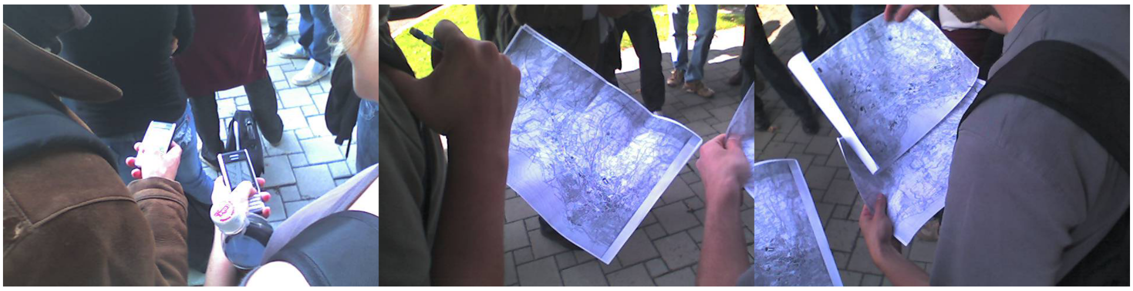


Figure 6-7: Settings of the initial walk conducted during the second field study. (Left) Students sharing advice on how to use the application. (Center and Right) Students annotating the paper maps used during the walk

I began meeting the students during one of the first lectures of the course in October 2007, where I could present the application and the basic functionalities. Each participating student received a phone and one-to-one instructions on how to start the application and get started with the annotations. This session lasted about 2 hours. In the same day, students were invited to the first collective walk in the west of Lausanne. The teachers of the class chose a particular path crossing many of the areas of this region that were of interest for the class. During this walk, students had the possibility of testing for the first time STAMPS. I participated in this initial activity to answer questions, solve problems and observe spontaneous forms of usage (see left frame of figure 6-7). Later on, during the rest of the course, I met regularly the students to make sure they were not experiencing problems with the application and to understand how they were using it. The course lasted almost three months and the students had the phone with the installed

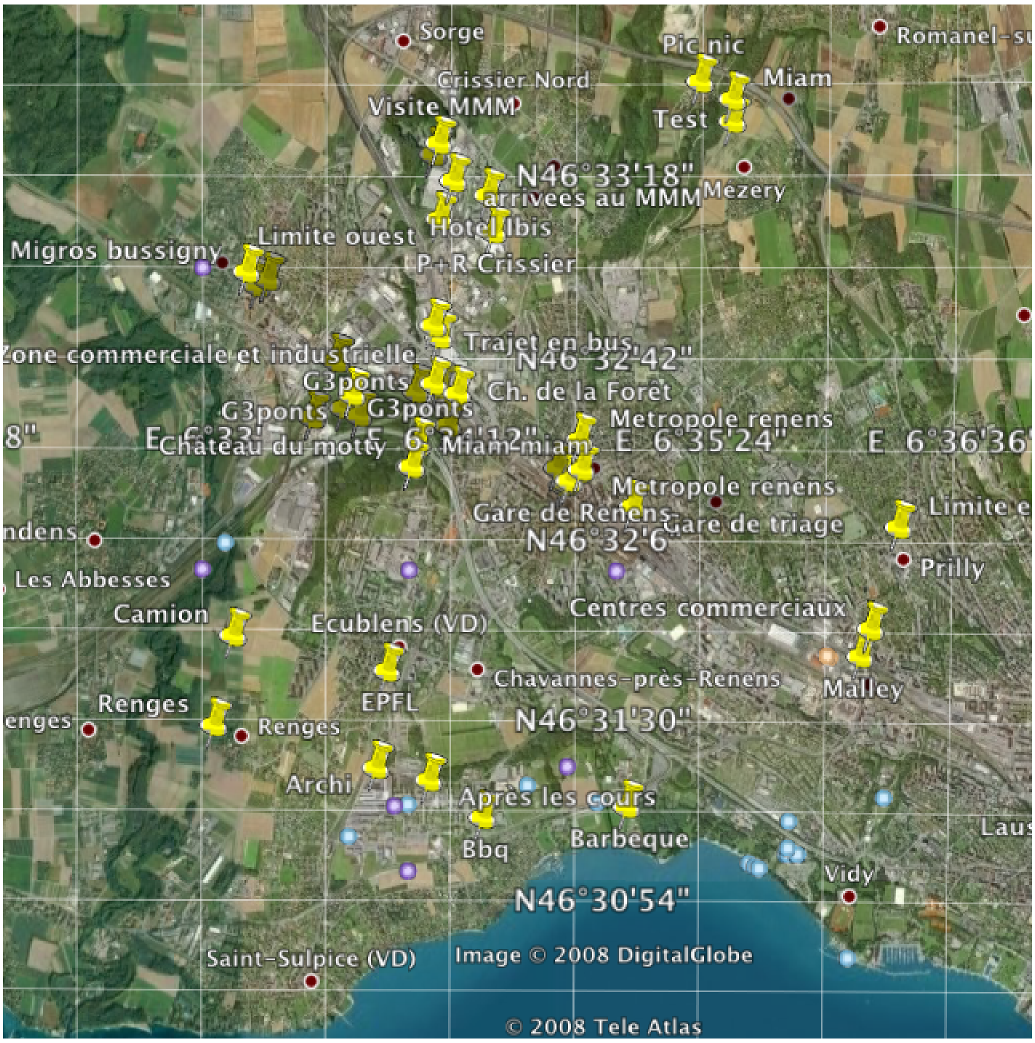


Figure 6-8: Annotation area of the second field trial. Pinpoints represent messages (© Google Inc.)

STAMPS application at their disposal during this period. As the end of the course and as a last step, I invited the students to a final interview, where I asked specific questions on the way the produced annotations have been used for the final report of the class. Unfortunately, I managed to interview only four participants (they all belonged to the 'curious' group, as explained below and were some of the most active in the trial). The others declined invitations to discuss the experiment further.

6.3.2 Results

Students created 50 map annotations during the trial (see figure 6-8), none of which was a reply to an existing message. However during the three months of the trial, 300 log files were collected, accounting for about 28 hours of login time. This time, the users were grouped under two profiles only, as I could not find participants using the system with the same frequency of the most actives participant of the first trial. These two groups reflect the same distinctions highlighted in the previous study: *the passives* (8 participants) and *the curious* (8 participants). Curious produced at least two messages (min. 2, median 2, max. 13 messages) and logged in the system at least four times (min. 4, median 15, max. 28 times), whereas passives did not produce any message and logged in only few times (min. 2, median 8.5, max. 64 times). Table 6.5 reports some quantitative statistic of use of the system during the trial. Interestingly, passives made longer sessions (min. 1.2, median 20.6, max. 64.1 minutes) compared to curious (min. 1.4, median 13.7, max. 88.7 minutes) and both curious and passives of experiment 2 stayed logged in for a longer period of time compared to the people of the same groups in the first experiment (C: 13.68 minutes, P: 20.57 minutes).

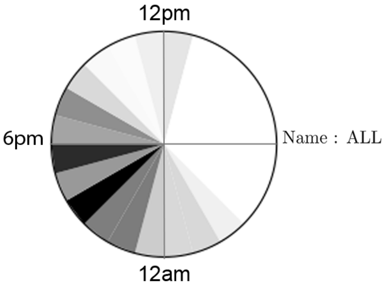
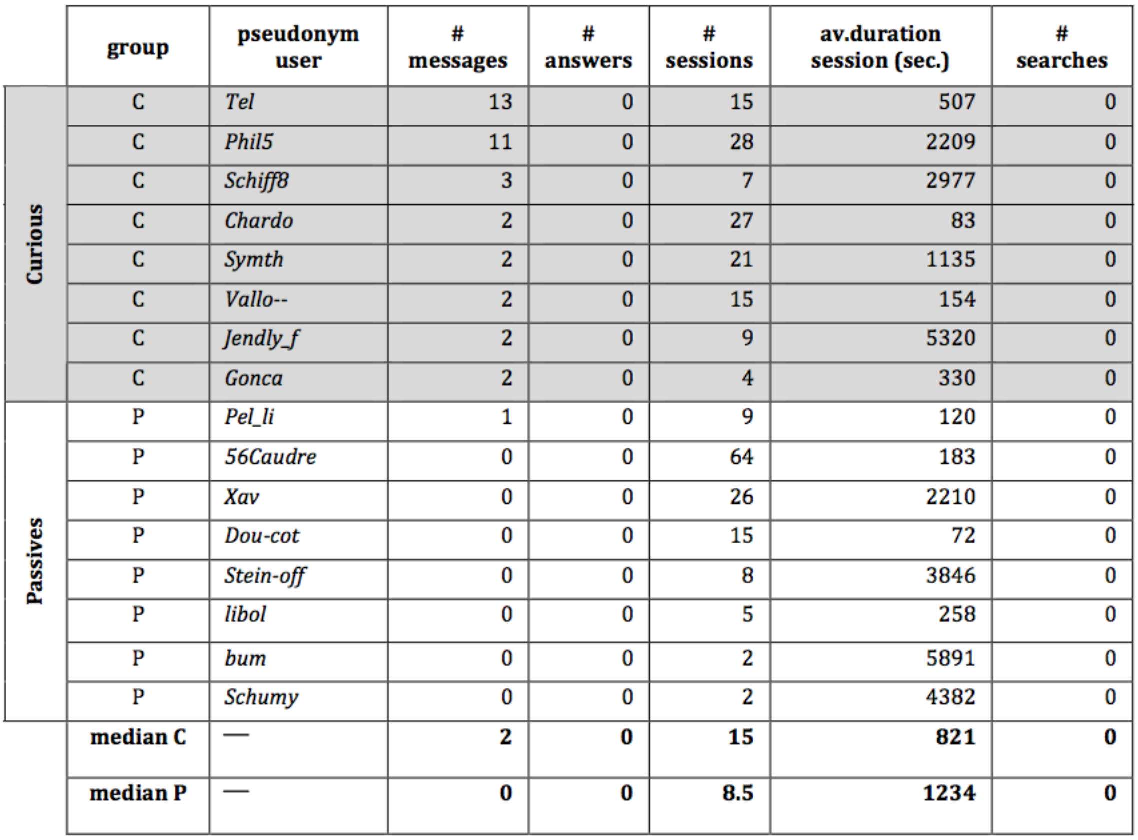


Figure 6-9: Cumulative representation of the login hours for all the participants of the second experiment

Additionally participants of this group used the system prevalently during working hours, the peak was at 4pm, right after the end of classes, while participants of the first experiment logged in during commuting time (see figure 6-9). Besides these rough differences, I could not observe any qualitative difference in the browsing style that participants in these two groups adopted. Movements over the map seemed comparable to those of the 'adopters', as observed in the previous experiment.

During the three months of the field trial, participants did not run any queries in the database. This is somewhat obvious giving the low number of messages available in the system and the low level of participation. Therefore, messages were mainly retrieved by browsing the map. Another interesting point was that contrary to the way messages were posted in the Geneva experiment, students in Lausanne posted mainly messages concerning a certain part of the city while being physically close to the point they were talking about. This argument was derived conducting the same analysis explained in section 6.2.3. The chart reported in figure 6-10 supports these findings as it shows that the average distance between the anchor points of messages posted from

Table 6.5: General statistics for three months of system usage of the second experiment



	group	pseudonym user	# messages	# answers	# sessions	av.duration session (sec.)	# searches
Curious	C	*Tel*	13	0	15	507	0
	C	*Phil5*	11	0	28	2209	0
	C	*Schiff8*	3	0	7	2977	0
	C	*Chardo*	2	0	27	83	0
	C	*Symth*	2	0	21	1135	0
	C	*Vallo--*	2	0	15	154	0
	C	*Jendly_f*	2	0	9	5320	0
	C	*Gonca*	2	0	4	330	0
Passives	P	*Pel_li*	1	0	9	120	0
	P	*56Caudre*	0	0	64	183	0
	P	*Xav*	0	0	26	2210	0
	P	*Dou-cot*	0	0	15	72	0
	P	*Stein-off*	0	0	8	3846	0
	P	*libol*	0	0	5	258	0
	P	*bum*	0	0	2	5891	0
	P	*Schumy*	0	0	2	4382	0
	median C	—	**2**	**0**	**15**	**821**	**0**
	median P	—	**0**	**0**	**8.5**	**1234**	**0**

the same antenna is below 500 meters, which is two times more the average diameter of a GMS cell in urban areas. It must be noted that this technique analyzes only those cellular antennas for which I had at least two records of messages posted while the mobile phone was connected to them.

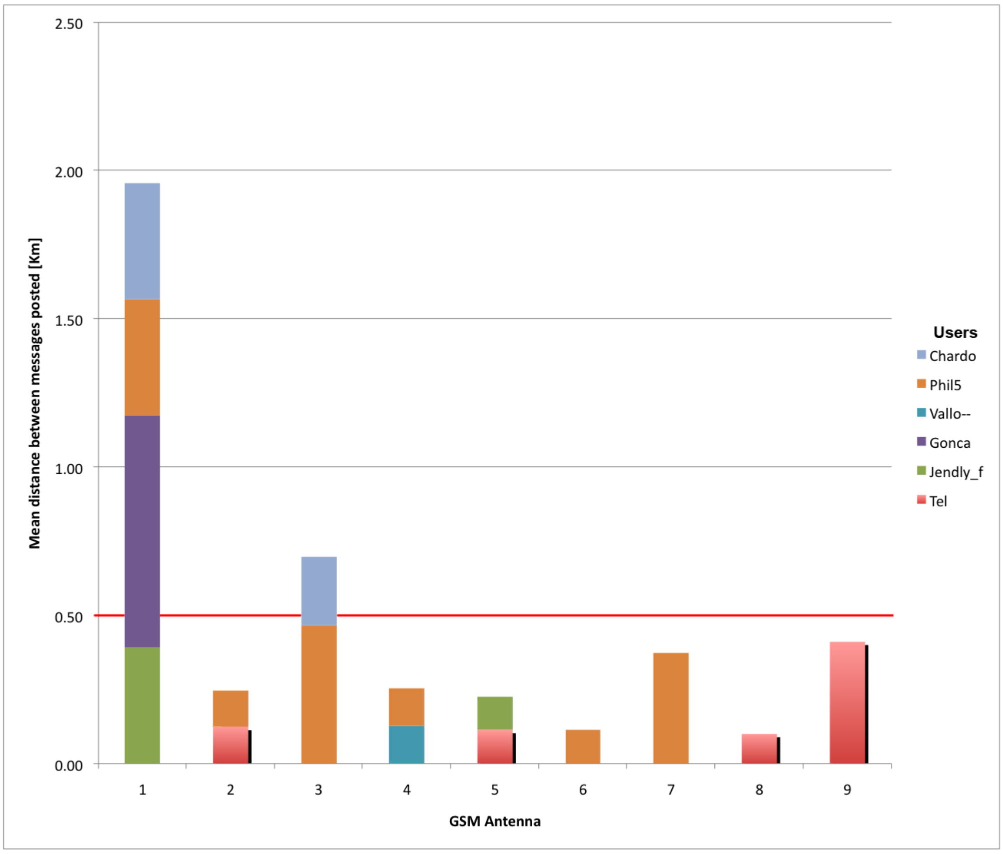


Figure 6-10: Average distance of the messages posted while connected to the same GSM antenna in the second experiment. The colors represent the proportion of message posted by each participant using that particular antenna

Next, the messages posted in the system were analyzed and I observed that it was not possible to apply the coding scheme used in the previous experiment (described in section 6.2.3). In fact, students wrote sentences that were not directed to other students or participants but to themselves. This argument was derived from the way they framed the title and the content of their messages. For instance, the user Tel posted three messages in sequence while conducting a field observation. The three messages were titled: "*visit MMM*", which is not self-explanatory. This was probably a technical name given to a particular building that she had to report on. Additionally, the content of the messages were detailing some specifics of the buildings: e.g., "*there isn't a pedestrian access on this side*", and "*small trail descending to the sorge*" (*La Sorge* is a river crossing the area under study). Notes contained also quantitative annotations for their report.

For instance, another student annotated the time required to walk to the bus stop from one of the villages' center: *"Stop bus 7, about 8min from Prilly"*. As a final example, another student took notes to delimitate their area of interest: *"Est limit: est limit of our zone of study, roundabout center Prilly"*. Of course with some effort, most of the details of these messages can be understood from an external observer but they were not meant for a general public when they had been written. The reader would be right to object that this style of writing was somewhat influenced by the course assignment. However, it must be noted that by that assignment students were asked to produce annotations for their group and for their entire class. However, as interviewed students declared, other people's notes were difficult to understand even for students belonging to the same working group.

As a final step, the content of the interviews was analyzed. I asked the students how they used the annotations taken with STAMPS in their final report. During the period, only four students downloaded their annotations from the web site (Tel, Phil5, Schiff8, and Symth). All of them said that they looked at the annotations only a couple of times but that they did not contain any fundamental information for their report. During the period, they kept using paper annotations (see center and right frame of figure 6-7). Interviewed students claimed that it was faster to take annotations on paper than going through a series of 'clicks' in STAMPS: *"It's more practical. I am pretty used to working with paper maps!"* (Phil5). The interviewed students agreed that a mobile annotation tool like STAMPS might be beneficial for their work, however they all said that such tool should allow easier note taking: *"it should be easier to record a message ... maybe it can record my voice"* (Tel). Also the availability of an automatic location acquisition was mentioned repeatedly during the meetings: *"If it could track my position automatically, like a GPS, that would save some of my time"* (Symth).

6.3.3 Discussion

I developed this second experiment to answer the following research question: *Q4, do the results of the first three research questions change when the participants use the CAS tool in a structured task?* The weak results presented in this section suggest that indeed, this was the case. The analysis of the logs shows peculiarities of use of STAMPS by the urban planners. Communication was self-directed. Annotations were only understandable by the author of the note. These notes were produced during working hours, a different result compared to the previous study. Clearly the 'sharing' functionality of the messages offered by STAMPS was less important in this scenario. One possible explanation was that students usually walked the area together with their classmates. They could share directly, and more efficiently, their impressions via voice because they were face-to-face.

Students did not use annotations extensively. The produced annotations were minimal. While the majority of the annotations produced in the first study were produced while on the bus or the metro, the notes produced in this study were taken at the actual place where to which they referred, thus suggesting that the content of the note was not known to the author beforehand. Additionally, few students downloaded the notes taken during the period from the web site (those who did belonged to the curious group). Answers provided from the interviews suggested that the electronic notes constituted only a fragment of the information necessary for the course and therefore were not of interest for most of the students that took them.

An important point that emerged during the interviews was that some students took mainly notes using the paper-based support instead of using the experimental tool (see central and right frame of figure 6-7, at page 175). This support has still advantages over STAMPS as notes could be taken faster on paper. However, paper maps also have drawbacks: a paper map presents information with a fixed projection scale (e.g., one cannot zoom to a certain part of it). Subsequently, many notes written over the same area can result illegible by the same author later on, or by another person.

An important critique of STAMPS that emerged during the interviews was that positioning is manual, therefore it is subject to mistakes and it requires interaction time to define the right section of the map. Perhaps, with an automatic positioning mechanism and with faster inputting methods, the gap of this application with a low tech material as the paper could be filled and the advantages of an electronic form of annotation could have a major impact on this particular activity. I must say that developing lighter or more intuitive inputting mechanisms was not the objective of this work. Wang and Canny (2006) conducted a comparative user study to this end and found that a technique combining a quick photo capture on the place and an offline editing was the most favorite method in ease of use for producing location-based annotations. Finally, I think STAMPS could help urban planners during their note-taking activity, but only if the cost of inputting notes is less than that required for the same annotations to be taken on paper. This was not the case with the current prototype.

6.4 Conclusions

Both studies led to very disappointing results, however this can be considered, per se, an interesting result. What these field trials highlighted is that having a nice set of features is not enough for an application to be adopted by a group of users. More research is needed to understand whether ubiquitous collaborative annotations of maps are useful and how best to support this activity. The studies reported in this chapter should serve as a cautionary tale to researchers who are trying to build such systems. The social characteristics of the annotation activity can be as significant as the design of the interaction mechanism to the user in its success or failure.

The reasons of failure of the first field trial stem from the lack of a critical mass of users, the lack of useful content, and the limited social awareness (the difficulty in seeing who posted what and the lack of social networks in study). One of the contributions of this work was to identifying the critical mass necessary for the success of such application not just with a number of participants, but rather with a group of peers, with whom sharing localized resources might have had an important value for the participant.

This first study offers an interesting perspective on the kind of messages that users would annotate a map with. The combined results of the first and second study showed that the range of possible categories of location-based annotations is related to the communicative goals of the emitters of these messages. While users of the first field trial targeted the whole population and produced mainly information on local resources, users in Lausanne did the opposite, producing self-directed messages that were difficult to understand outside of the scope of the course for which they were written.

More interestingly (and importantly for this thesis), the first study shed some light on the relation between messages and physical locations and the way a CAS application can mediate this relationship. STAMPS was designed around a 'map-first' interaction paradigm. The content of the system was organized geographically by the map, and therefore users accessed content mainly from the map and not the other way around. Participants used the map to restrict the region of interest and then retrieved the annotations available in that particular region. This mechanism perhaps had consequences also on the way messages were produced. In fact, STAMPS was seen as an asynchronous communication device even if, in principle, it could support synchronous communication. Therefore, this result suggest that this specific interface features might have a key impact on collaboration outcomes.

A more controlled experiment is therefore required to verify this hypothesis. Understanding more precisely, and quantitatively, how the organization criterion of the messages affect users' performance in a collaborative task can help CAS designers in develop better support for collaboration. This is one of the objectives of the next chapter.

A final note concerns the broader lessons that I learned from the experience. Both field trials provided me with ideas on how to improve the design of such research. First of all, the time span of these trials was too long. If I were to run a similar observation again, I would reduce the length of the period and include iterations: testing sets of design features in shorter periods of time, taking time to analyze the results and design possible improvements, and run other sets of observations. A second point that I now consider important in the observation of mobile applications such as STAMPS is the involvement of a large group of friends. During the trial, I could observe the importance for participants to be connected with their group of peers. The key of social platforms is to allow users to express their personality to their close acquaintances.

Chapter 7

The Effects of Explicit Referencing in Distance Problem Solving Over Shared Maps

This chapter describes the controlled experiment that was conducted to answer the fifth research question. It reports a quantitative analysis that was conducted to asses the effect of Explicit Referencing on collaborative work at a distance. While the previous experiments have been organized with mobile devices on the field, this experiment was developed in laboratory with fixed workstations and using eye-tracking displays.

7.1 Introduction

The introduction of this thesis provided examples of how in certain situations collaborators are required to coordinate their actions with a limited bandwidth available. In these situations, people might use Collaborative Annotation Systems, which implement the *Explicit Referencing* (ER) mechanism introduced in chapter 3. As discussed in the previous chapters, different designs of this mechanism can impact on the communication flow. Chapter 4 analyzed a number of systems supporting ER, introducing a framework to differentiate and compare their features. One of the main dimensions of this scheme is the way messages are organized. Two solutions are possible: a time-based criterion (messages are represented following their temporal order of emission) or a context-based criterion (messages are represented according to the places of the map to which they refer). Qualitative observations reported in chapter 6 support the idea that interfaces built around the latter criterion might influence collaborative interactions that people have in such systems. Participants of the field trials reported in the previous chapter, who used an ubiquitous-CAS in which messages were organized by the map, spontaneously adopted asynchronous communication strategies. Therefore, this design might have negative consequences in collaborative situations where synchronous communication is required. More evidences are required to verify this hypothesis. Therefore, I chose to consider the message organization criterion as an experimental factor in the quantitative analysis presented here.

Moreover, the hypothesis suggested by the literature discussed in chapter 3 is that enabling explicit connections of utterances to the shared workspace helps in disambiguating references to shared objects, thus improving the collaboration process. However, if utterances are overlaid on a map (messages organized by context), they are no longer sequentially displayed as in a chat window. This visual dispersion of utterances may be detrimental to the joint construction of the common ground. Both the conversation and the shared workspace belong to the collaboration context, the general frame of reference that is used by pairs to make sense of each other's intentions. However for the sake of this study, I will distinguish between the conversational space, or *conversation-context*, and the support provided by the shared workspace, which I will name *task-context*.

From the literature review discussed in the previous chapters, it is clear that Explicit Referencing has an influence on the communication processes that collaborators might employ while solving a task collaboratively. Still, little is known about the interplay between these two communication contexts. Are they equally important in achieving the goals of the task? In this chapter, I investigate whether supporting ER, in a task involving spatial coordination, enhances collaboration. This chapter addresses the fifth research question of this thesis: *Q5, does the availability of Explicit Referencing enhance the performance in a collaborative problem solving task at a distance?* This generate 3 hypotheses for this this empirical study:

H1 Explicit referencing leads to better team performance;

H2 CAS organizing messages according to the position on the map to which they refer lead to inferior performance;

H3 Explicit referencing makes communication more efficient (fewer sentences, with fewer words);

This chapter addresses the following trade-off: supporting references to the workspace (task-context) and maintaining the clarity of the conversational space (conversation-context), in the situation in which collaborators have to perform a joint task while being not co-located. I report the results of an experimental study where I compared performance and processes of teams who had to organize a music festival on the EPFL campus (see next section). They used a chat tool and a shared map (see section 7.2.6). I compared experimental conditions where participants shared information linked to the map against a control setup where participants used a standard chat tool, as explained in the following section.

7.2 Method

Participant pairs had to collaboratively perform the following task: organize a festival on their university campus, collaborating remotely using a chat tool. Completing the task required deciding which parking lots would be used by the festival attendants, where to position the three stages of the event, and how to allocate six artists to the three available stages. They therefore had to perform a number of optimizations, such as minimizing the distance between the chosen parking areas to the initial stage and between stages according to the schedule of the events. Additionally each parking lot had a different rental price that was somewhat proportional to its capacity. One of the constraints required the subjects to minimize the budget for the concert. Finally, as setting up a concert on a stage required appropriate "sound checks", subjects had to choose the order of the concerts so as to minimize the waiting time of the spectators, and an appropriate distance among the stages so as to minimize the disturbance of sound checks on concerts already in progress. To summarize, four goals were presented to the participants (see the full instruction sheet reported in appendix A, on page 327):

1. to minimize the distance the participants will have to walk to reach the stages;

2. to maximize the distance between the stages so to avoid audio disturbance;

3. to minimize the expenses for renting the parking lots;

4. to decide the schedule of the concert reducing the overlap of the events on the same stage and minimizing the participant' s walking distance to move around between the stages.

The subjects had to position a series of icons on a campus map: a number of 'P' signs to mark the active parking lots, three stage icons and six small circled numbers, one for each event to be allocated (part (b) of figure 7-3 and figure 7-5). The positions of these icons were not synchronized across the participants' displays: a subject could not see where the other would position her icons. This task was artificially made complex (e.g., not WYSIWIS) so as to augment the difficulty in finely positioning the icons between the two screens and so that I could observe how arising conflicts could be solved at a linguistic level and/or with different communication tools. This design was also chosen in order to separate the effect of the *feedthrough*, as explained in section 3.3.2, at page 70, and the availability of a shared display on the team's performance, from that of Explicit Referencing, the focus of this research.

7.2.1 Participants

Hundred-and-twenty students (55 women and 65 men, mean age = 23.5 years, sd = 1.2 years) of the Swiss Federal Institute of Technology in Lausanne volunteered to participate to the experiment. They were selected based on their mother language, their course year, their faculty and their knowledge and use of computers and, in particular, chat applications. All volunteers were native French speakers. I did not recruit participants in the first or the second year of their program as this could effect the level of their knowledge of the campus site and, in turn, on the task performance.

The subjects did not know each other and were randomly matched from different faculties. Students from Architecture or Civil Engineering were excluded as they could have biased the results as they are more used to working with maps. They were recruited using an e-mail call for participation and a short telephone interview, which helped to ascertain that they regularly used a chat application and that they did not have any ocular disabilities (e.g., colorblindness). Each participant was remunerated 30 Swiss Francs ($\sim$18.30 EUR, or $\sim$24.85 USD).

Participants were randomly assigned to 60 dyads. Fifteen dyads were assigned to each of the four conditions described in section 7.2.5.

7.2.2 Apparatus

The members of a pair were each seated in front of identical desktop computers with 17-inch LCD eye-tracker displays (maker: Tobii, model: 1750, now called MyTobii D10), and located in two different rooms (see figure 7-1). The settings of the rooms, the working table and the light conditions were identical. Particularly, I also partitioned off the study space using shelves to reduce distractions from other objects present in the room.

Participants sat unrestrained approximately 60 cm ($\sim$ 24 inches) from the screen. The tracker captured the position of both eyes every 20 ms. The participants went through a 5-point calibration routine.

7.2.3 Procedure

On arrival, participants were each given an instruction sheet (reported in appendix A, at page 327) containing the rules they had to respect in placing the elements on the map, information on how to evaluate their solution, and the principles behind the calculation of the score. After, they were asked to watch a short video summarizing the paper instructions and explaining the particular communication tool they were to use to collaborate. Prior to starting the task, the participants could ask questions to the experimenter if they had any doubts about the video or written instructions they were given.

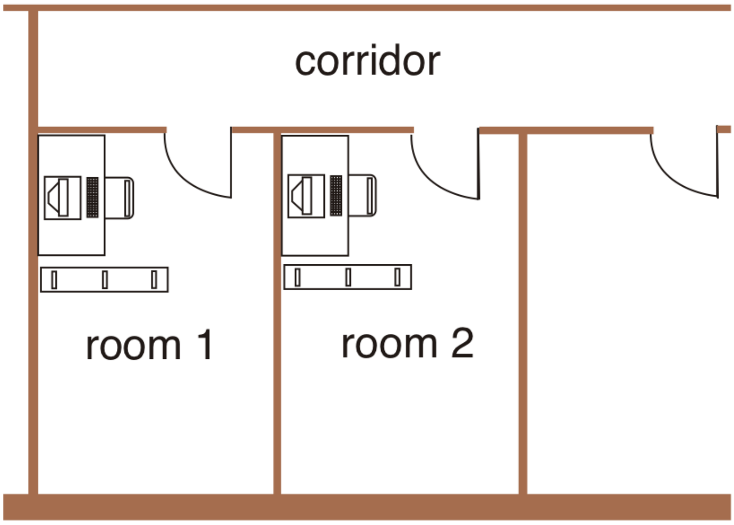
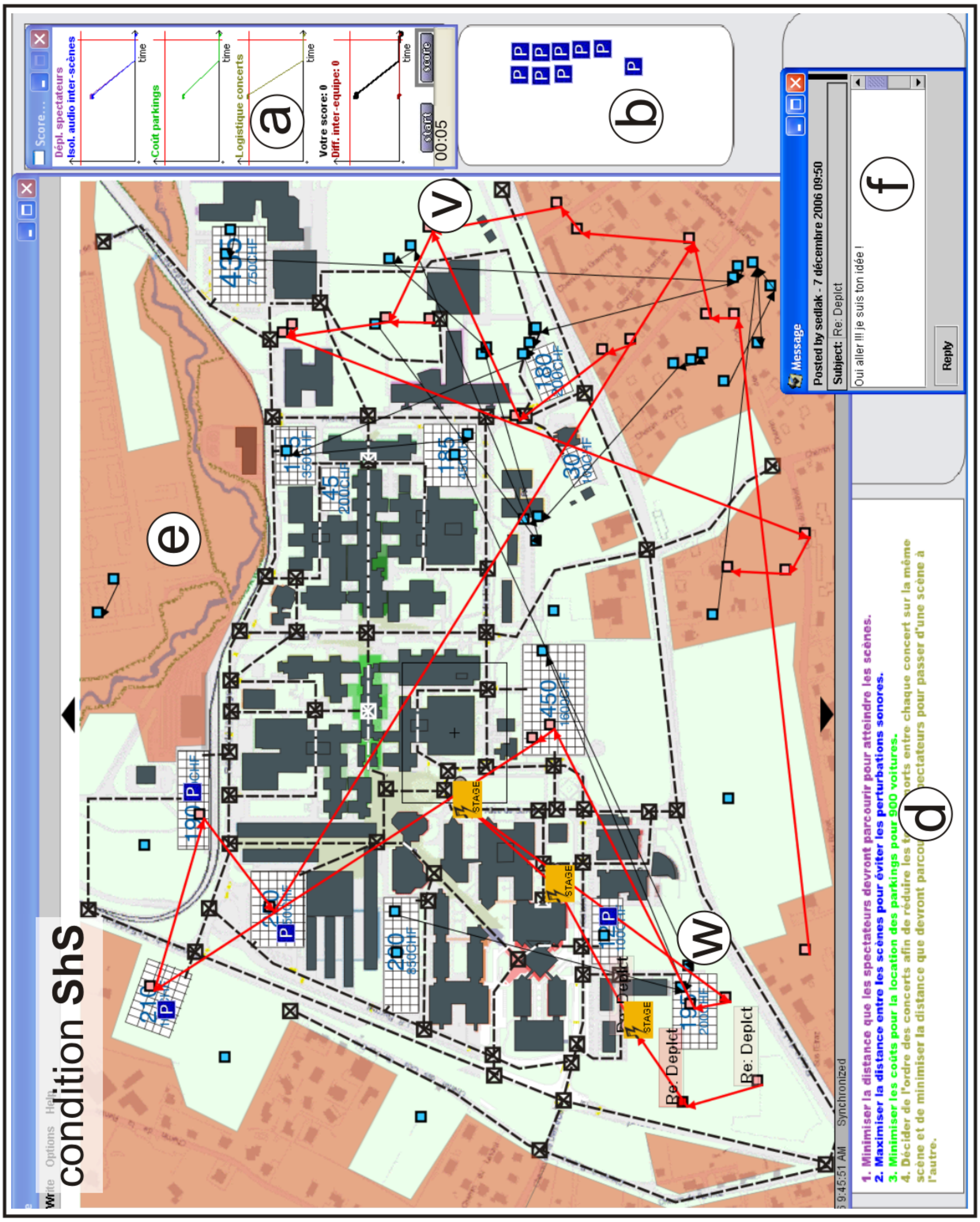


Figure 7-1: Setting of the experiments. Each participant was seated in a different room and used an eye-tracking screen. I partitioned the room space using shelves to reduce distractions from other objects present in the room

During the task, each participant had at her disposal: a feedback tool (part (a) of figure 7-3), a map of the campus (part (e) of figure 1) and a chat application to communicate with the partner. The feedback tool offered a score button (part (v) of figure 7-3), to display of a number between 0 and 100. This score was computed by comparing the proposed solution with the optimal solution that was calculated once for all the experiments. This tool also presented four graphs that would display four sub-scores one for each goal and the combined team-score. Each graph presented a horizontal red line, representing the maximum score that could be achieved with the given constraints and a vertical red line marking the time limit of the task. The tool also showed the remaining time to complete the task in the bottom-left corner (part (v) of figure 7-3). This tool also kept a detailed log of the users' actions (see an excerpt in appendix B, at page 331)

The task lasted 45 minutes. As the task required multiple optimizations, I allowed each pair to submit multiple solutions to solve the task, ultimately selecting the best score for each team. Pairs were instructed to find the configuration leading to the highest score and to follow a collaborative paradigm. In fact, the pairs were warned that every time they pressed the score button, the system checked the position of the icons on the two machines. Pairs were told that the number of differences found was detracted from the obtained score. They were also advised to take advantage of the feedback tool and the available time to test the maximum number of different configurations.

At the end of the experiment, the participants were invited to participate in a debriefing session where they could ask questions and discuss the outcomes of the experiment. I conducted the interviews, asking specific questions on their interaction and to record the answers given. In



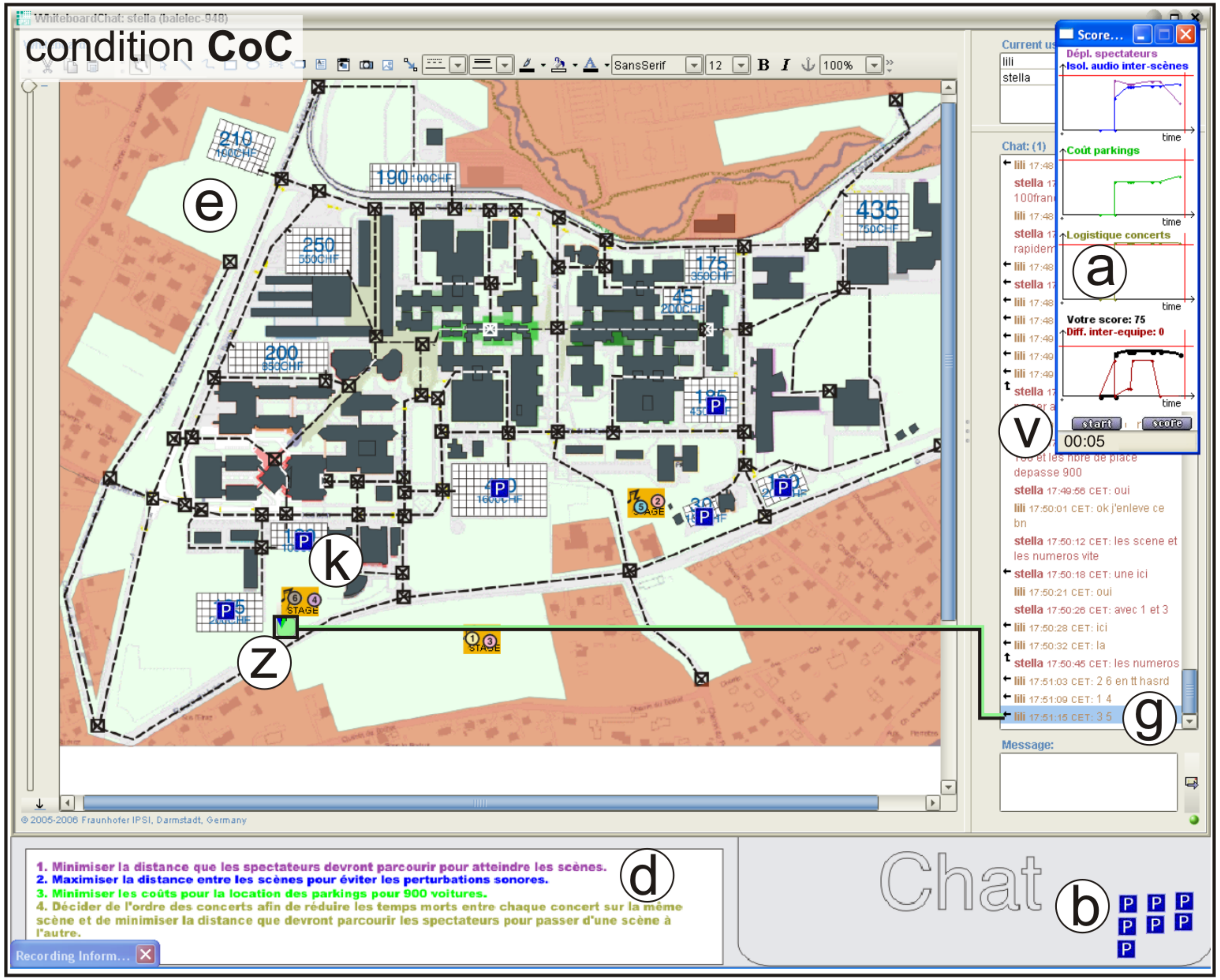
Posted by sedlak - 7 décembre 2006 09:50

Subject: Re: Deplct

Oui aller !!! je suis ton idée !

1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les transports entre chaque concert sur la même scène et de minimiser la distance que devront parcourir les spectateurs pour passer d'une scène à l'autre.

Figure 7-2: Experiment setup in the ShoutSpace condition (ER – noHist): (a) feedback tool; (b) icons used during the task; (d) reminder of the task goals; (e) map window; (f) ShoutSpace chat message window; (v) score button and countdown timer; (w) in ShoutSpace message anchors are small squares on the map. Clicking on a square opens the message window f. In ShoutSpace answers to existing messages are displayed with an arrow connecting the child with the parent



depasse 900

stella 17:49:56 CET: oui

lili 17:50:01 CET: ok j'enleve ce bn

stella 17:50:12 CET: les scene et les numeros vite

stella 17:50:18 CET: une ici

lili 17:50:21 CET: oui

stella 17:50:26 CET: avec 1 et 3

lili 17:50:28 CET: ici

lili 17:50:32 CET: la

stella 17:50:45 CET: les numeros

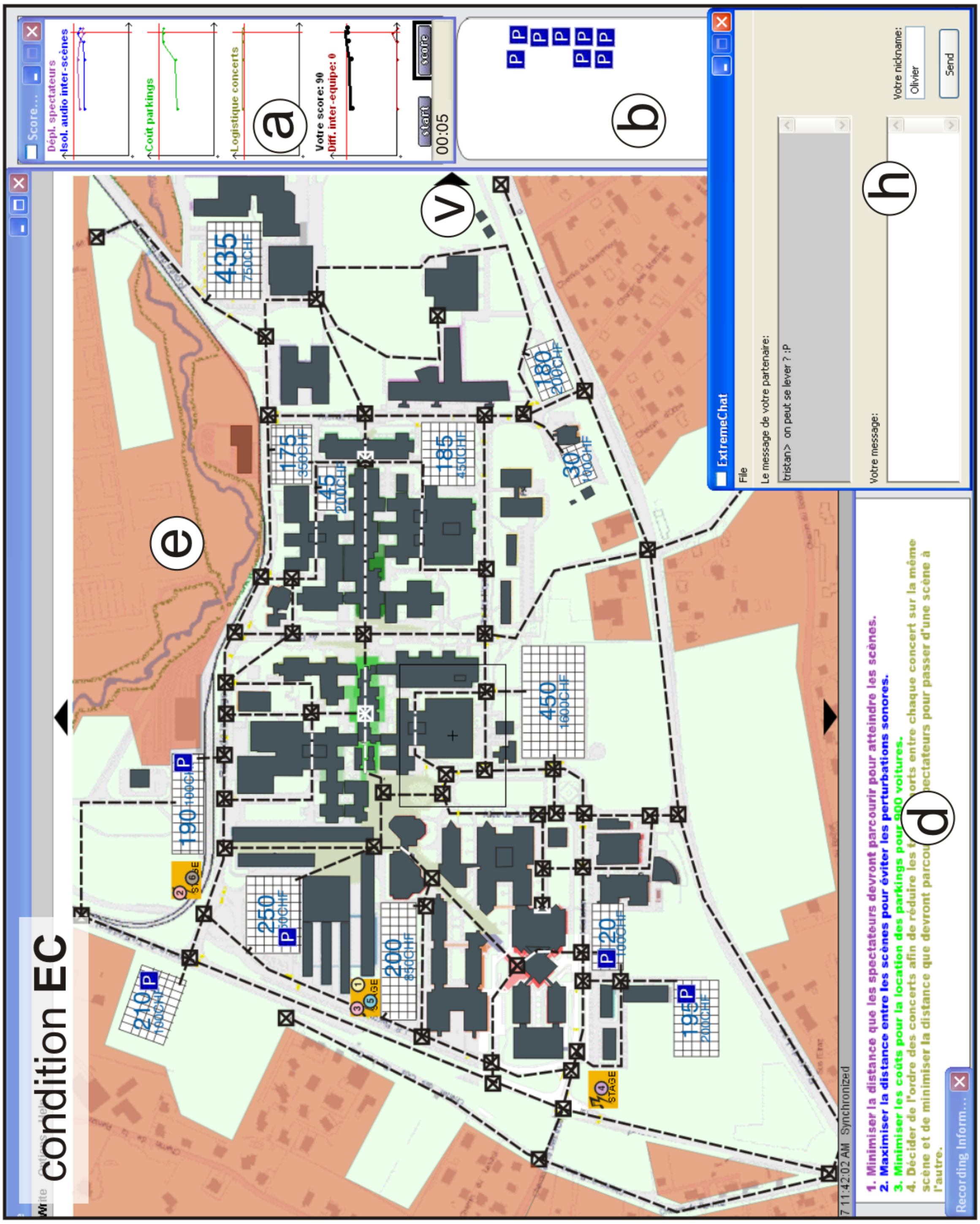
lili 17:51:03 CET: 2 6 en tt hasrd

lili 17:51:09 CET: 1 4

lili 17:51:15 CET: 3 5

1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les temps morts entre chaque concert sur la même scène et de minimiser la distance que devront parcourir les spectateurs pour passer d'une scène à l'autre.

Figure 7-3: Experiment setup in the ConcertChat condition (ER – Hist): (a) feedback tool; (b) icons used during the task; (d) reminder of the task goals; (e) map window; (g) ConcertChat chat message window; (k) example of how a stage icon is positioned and two concerts assigned to that stage; (v) score button and countdown timer; (z) in ConcertChat it is possible to connect the message window to a point of reference with an arrow



7 11:42:02 AM Synchronized

1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les t orts entre chaque concert sur la même scène et de minimiser la distance que devront parcou pectateurs pour passer d'une scène à l'autre.

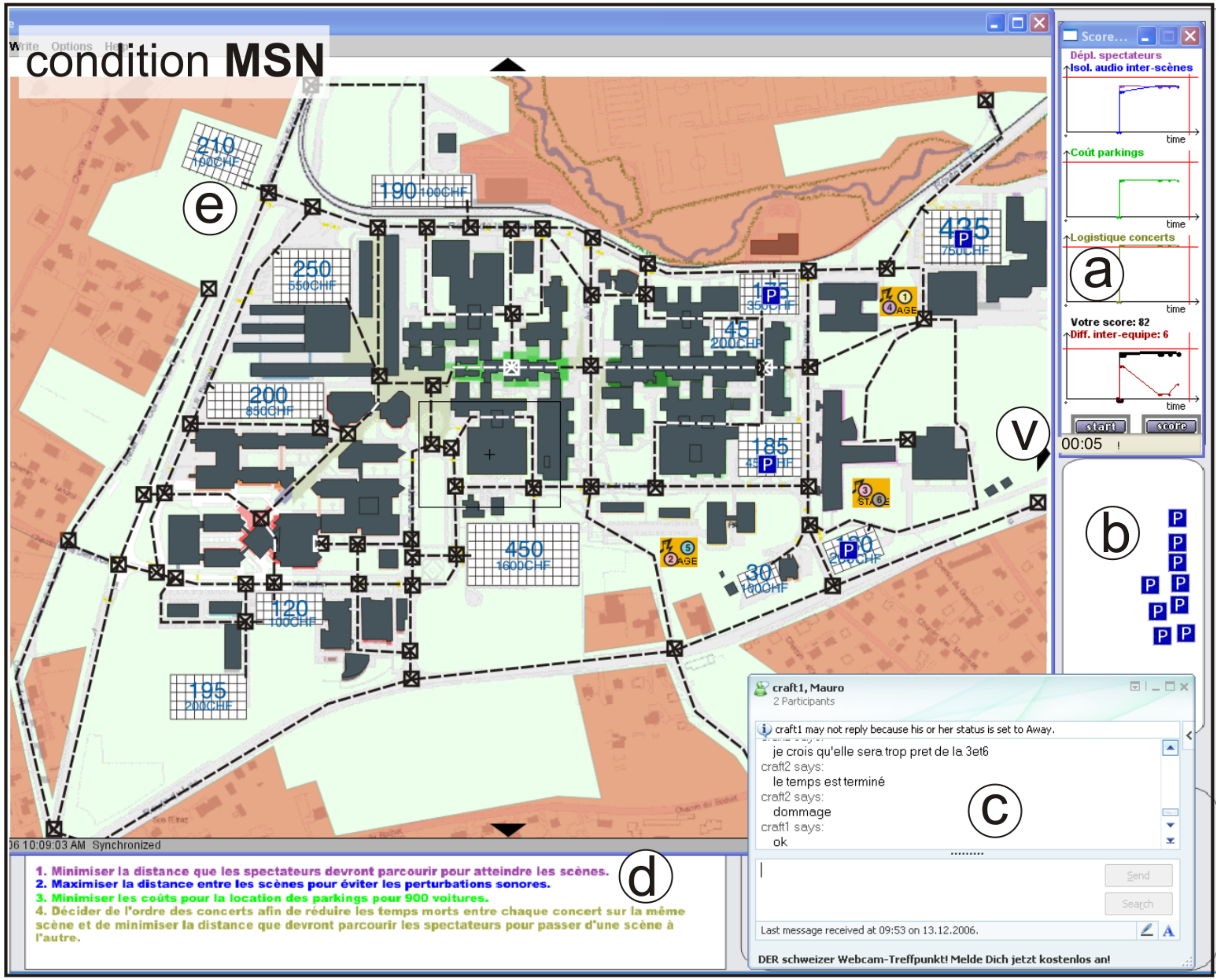
Le message de votre partenaire:

tristan> on peut se lever ? :P

Votre message:

Votre nickname: Olivier

Figure 7-4: Experiment setup in the ExtemeChat condition (noER – noHist): (a) feedback tool; (b) icons used during the task; (d) reminder of the task goals; (e) map window; (h) ExtremeChat message window: only the last message of the partner is visible to the participant; (k) example of how a stage icon is positioned and two concerts assigned to that stage; (v) score button and countdown timer



craft1 may not reply because his or her status is set to Away.

je crois qu'elle sera trop pret de la 3et6

craft2 says:
le temps est terminé

craft2 says:
dommage

craft1 says:
ok

Last message received at 09:53 on 13.12.2006.

DER schweizer Webcam-Treffpunkt! Melde Dich jetzt kostenlos an!

1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les temps morts entre chaque concert sur la même scène et de minimiser la distance que devront parcourir les spectateurs pour passer d'une scène à l'autre.

Figure 7-5: Experiment setup in the MSN chat condition (noER – Hist): (a) feedback tool; (b) icons used during the task; (c) MSN chat message window; (d) reminder of the task goals; (e) map window; (k) example of how a stage icon is positioned and two concerts assigned to that stage; (v) score button and countdown timer

particular, I used this opportunity to record qualitative information on the experiment. I asked whether the participants had any sort of conflict or misunderstanding during the interaction and what strategy they developed to position the icons at the same locations

7.2.4 Measures

The pairs were instructed to complete the task collaboratively, trying to minimize the number of mistakes in positioning between the icons on the two machines. I used the highest score achieved during the 45 minutes as the primary measure of task performance. At the beginning of the task, all the pairs were required to move the icons placeholders from the docking location (part (b) of figure 7-3) to the map. Pairs usually discussed an initial configuration, moved the icons, and pressed the score button. I used the time required to reach this initial configuration of the icons as a second measure of performance. As a third measure of performance, I measured the amount of solution space that was explored by the pair, as explained in section 7.3.1.

To understand how the pairs performed in different conditions, I explored several features of conversation structure: I looked at the number of words, number of utterances, and structure of turn taking. Then, I looked at the use of linguistic deictics (e.g. *"I want to use this parking lot"*) or other strategies of referencing to the shared map like labels (e.g., *"Let's take P200"*) and prepositional phrases (e.g., *"Place the stage below P450"*). I also recorded eye-gazes on the map to consider different strategies of map reading. I will detail the results of the eye-tracking analysis in the next chapter.

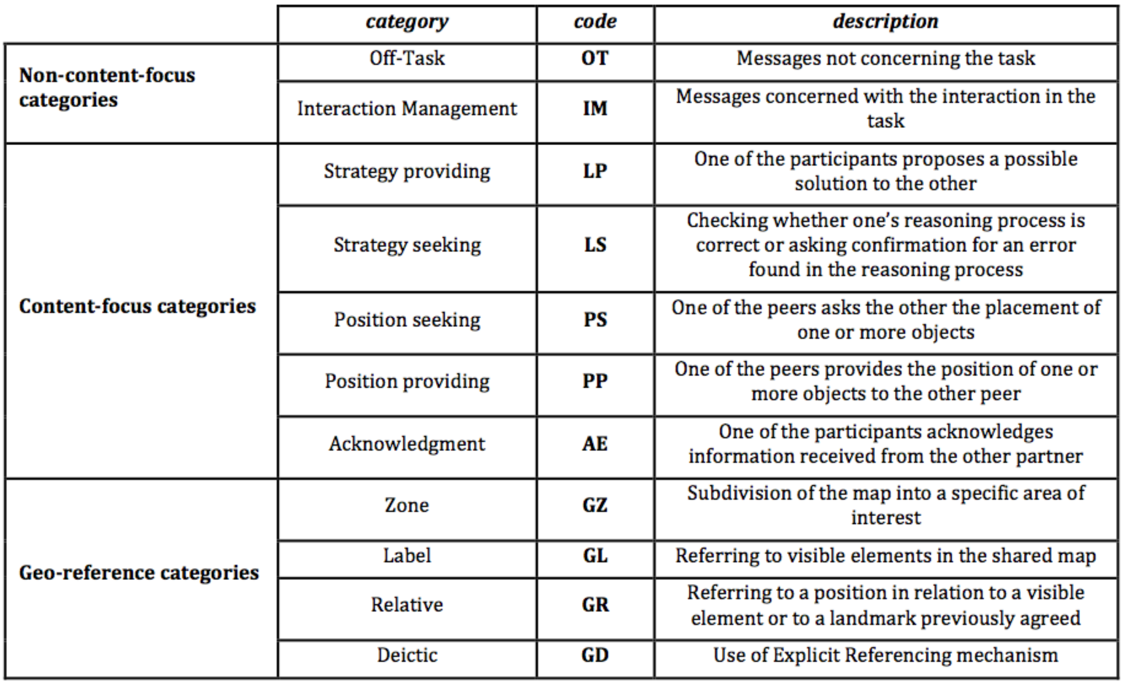
Additionally, I measured the mean time required in each trial to compose the messages exchanged. To achieve this goal, I calculated the time from the moment the user started typing the first letter of a message to the moment in which the message was sent to the collaborator. This calculation included the situations in which some characters were erased or entire parts of the message modified. This measure accounted better for effort required to produce a message than the raw number of characters of which the message was composed.

Finally, an average measure of the position mistakes of the icons was used to determine whether the pair followed a cooperative or a collaborative paradigm, as this could impact the results. A higher number of mistakes would have been a symptom of the pair's attempt to parallelize their effort to test multiple solutions.

Linguistic coding scheme

As the task required a fine regulation of the interaction, I manually coded the dialogue of each pair to account for differences across the experimental conditions. Although a purpose-generic coding scheme was suitable for my design, I developed my own coding scheme because the research question I asked required that only specific aspects of interactions be coded. The set of categories I used is based on three main distinctions: (a) between interaction that is "inside" the collaborative activity or "outside" (off-task activity) of it, (b) within "inside" activity, between interaction that is "content-focused" or "not content-focused" (social relation, interaction management, task management), and (c) within the "content-focused" interaction, between "strategy" and "position". Strategy messages refer to the discussion of hypothetical positioning of the icons in relation to a specific tactic, while position messages discussed the refer to the actual position of icons on the map once the strategy was agreed upon. In addition, I further subdivided each of these two categories into "seeking" or "providing", to differentiate their pragmatic attitude (see table 7.1).

Table 7.1: Coding scheme used to tag the corpus



	category	*code*	*description*
Non-content-focus categories	Off-Task	**OT**	Messages not concerning the task
	Interaction Management	**IM**	Messages concerned with the interaction in the task
Content-focus categories	Strategy providing	**LP**	One of the participants proposes a possible solution to the other
	Strategy seeking	**LS**	Checking whether one's reasoning process is correct or asking confirmation for an error found in the reasoning process
	Position seeking	**PS**	One of the peers asks the other the placement of one or more objects
	Position providing	**PP**	One of the peers provides the position of one or more objects to the other peer
	Acknowledgment	**AE**	One of the participants acknowledges information received from the other partner
Geo-reference categories	Zone	**GZ**	Subdivision of the map into a specific area of interest
	Label	**GL**	Referring to visible elements in the shared map
	Relative	**GR**	Referring to a position in relation to a visible element or to a landmark previously agreed
	Deictic	**GD**	Use of Explicit Referencing mechanism

A second code was associated to each message containing a reference to a specific position on the shared map (see bottom of table 7.1). Looking at the different strategies used by the participants across conditions, I found four recurring situations in which participants used references.

GZ While discussing a general strategy for completing a task, participants sometimes subdivided the map in sub-regions so as to handle a smaller number of features (e.g., "*Let's begin on the North-West*");

GL When choosing parking lots, for instance, they often referred to the label marking the lot's capacity, as this was a unique number on the shared map (e.g., "*Let's take P210*");

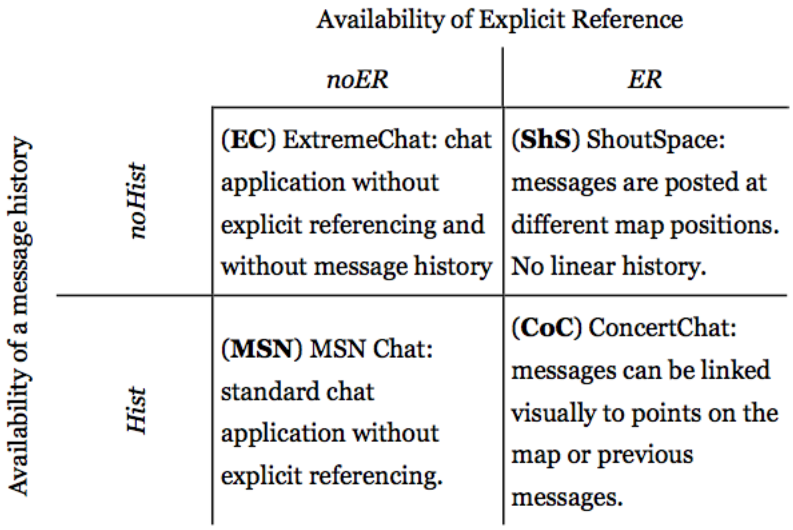
GR When participants had to position a stage, they often used previously established landmarks or visible features of the map to direct the positioning of the icons (e.g., "*On the right hand-side of P300*");

GD The last recurrent strategy used by the participants consisted of using the Explicit Referencing Mechanism used by the chat application in some of the experimental conditions (e.g., "*I suggest using this parking lot*").

7.2.5 Independent variables

My research question is to find out what is the impact of Explicit Referencing in collaborative problem solving at distance. I therefore varied the referencing support for the task-context (availability of ER) and for the conversation-context (the messages organization criterion): (1) can users relate an utterance to an element to the shared visual space (yes/no) (2) do users have access to a linear chat history (yes/no). The design was therefore a standard 2 × 2 factorial design, where Explicit Reference (*ER vs. noER*) and the presence of a linear conversational context (*Hist vs. noHist*) were between-subjects factors.

Table 7.2: Experimental plan of the quantitative experiment presented in this thesis



		Availability of Explicit Reference	
		noER	*ER*
Availability of a message history	*noHist*	**(EC)** ExtremeChat: chat application without explicit referencing and without message history	**(ShS)** ShoutSpace: messages are posted at different map positions. No linear history.
	Hist	**(MSN)** MSN Chat: standard chat application without explicit referencing.	**(CoC)** ConcertChat: messages can be linked visually to points on the map or previous messages.

7.2.6 Technical setup

Table 7.2 reports the four communication tools that I compared in the experimental plan. MICROSOFT MSN© is a standard chat application in which messages follow the temporal flow of the conversation (now called Microsoft Live chat[1]). ShoutSpace[2] and ExtremeChat are chat applications that were developed for this experiment, and ConcertChat was developed at Fraunhofer-IPSI, in Germany (Mühlpfordt & Wessner, 2005).

ExtremeChat (**EC**) is a rudimentary chat application that offers the persistence only of the last utterance emitted by the conversational partner. The next utterance emitted replaces the previous one (part (h) of figure 7-5). The organizing principle of the messages in EC does not follow a temporal criteria as newly emitted messages are not visible for the emitter and override older messages. Conversely, in EC messages are organized by space as they always appear in the same area of the interface.

ShoutSpace (**ShS**) allows the attaching of messages to a map: by default, the user sees only the anchor points of the messages (part (w) of figure 7-3). If she clicks on these, the messages appear in a pop-up window. Only one message at a time is visible to the user (part (f) of figure 7-3). In ShS messages are organized by the space of the map.

In ConcertChat (**CoC**), visual priority is given to the conversation. Connections to map locations are made by arrows connecting the message from the history panel to the map point (part (z) of figure 7-3), or to other messages in the history pane. Lines are refreshed as utterances move up the chat history (part (g) of figure 7-3).

Conditions **ShS** and **CoC** differ from condition **MSN** and **EC** in that they enable explicit referencing to a map-object, while conditions **MSN** and **CoC** differ from condition **ShS** or **EC** in that the chat history is displayed sequentially while in **ShS**, messages are scattered all over the map, and in **EC** is not possible to see messages older than the last partner's utterance and not even the user's own messages. Both **MSN** and **CoC** thus facilitate maintaining the conversation context and making implicit references. Additionally, **CoC** allows explicit references to previous messages in the chat history.

The explicit references created with **ShS** or **CoC** were part of the shared visual space and therefore synchronized on the two machines. However, the icons on each machine were handled by widget software that kept them on a topmost graphical layer. They were completely separated from the different communication tools tested in the experiment. Finally, the message input field in each of the four interfaces was about the same size.

7.3 Results

I will first describe the measured effects on the task resolution efficiency. Then, I will describe the differences in the process variables that I computed to account of the different resolution

[1] See `http://get.live.com/messenger/`, last retrieved March 2008.

[2] Several colleagues contributed to the development of ShoutSpace. See the acknowledgment section at the beginning of the thesis.

strategies. Section 7.4.1 will show how the variability of these measures can explain the obtained results. Finally section 7.3.6 will offer qualitative comparisons between the different conditions.

An alpha level of .05 was used for all statistical tests. All the measures presented refer to a sample of 15 pairs per condition. However, for those measures requiring manual coding I restricted the sample to 10 pairs per condition. These will be indicated case by case.

7.3.1 Task performance

This section presents findings concerning the first and second hypotheses: *H1, Explicit referencing leads to better team performance;* and *H2, CAS organizing messages according to the position on the map to which they refer lead to inferior performance;*. I employed three measures of resolution efficiency: the best score achieved by the pair during the resolution of the task; the time required for the pair to reach the initial positioning of the necessary icons on the map; and finally, a measure describing the variability of the different tested solutions.

Task score

Each time one of the participants pressed the "score" button of the feedback tool, the position of the icons on the two machines was recorded, and a score was computed comparing the proposed solution to the optimal solution for the given constraints. This score was a measure ranging between 40 points for a low-quality solution to 95 for the optimal solution. During the experiments the optimal arrangements of the icons was never found. However a pair found an almost optimal solution corresponding to a score of 94. The mean of best scores found across all trials was 83.

The availability of an Explicit Referencing mechanism had a negative impact on the score obtained by the teams. The pairs using a chat application supporting ER had lower scores compared to those in which this mechanism was not available (*ER*: m=81.63, sd=9.61 pts *vs. noER*: m=84.92, sd=4.07 pts; F[1,58]=13.37, p<.001). This result was not consistent with H1, which was predicting higher scores for trials with applications supporting ER. However, I should caution that this result was generated by averaging of the scores of ConcertChat, the highest of the experiment, with the scores of ShoutSpace, the lowest of the experiment.

The availability of a linear chat history had a large impact on the score achieved by the pairs. Pairs had substantially higher scores in the trials in which there was a chat application with a linear message history, compared to those in which messages where organized according to space (*Hist*: m=87.32, sd=5.08 pts *vs. noHist*: m=79.23, sd=7.41 pts; F[1,58]=81.22, p<.001). This result was consistent with H2, which was predicting lower scores for CAS organizing messages according to space.

The impact of Explicit Referencing on the score was inverted when a chat history was present. The Explicit Referencing × Message History interaction showed that ER was useful when the chat

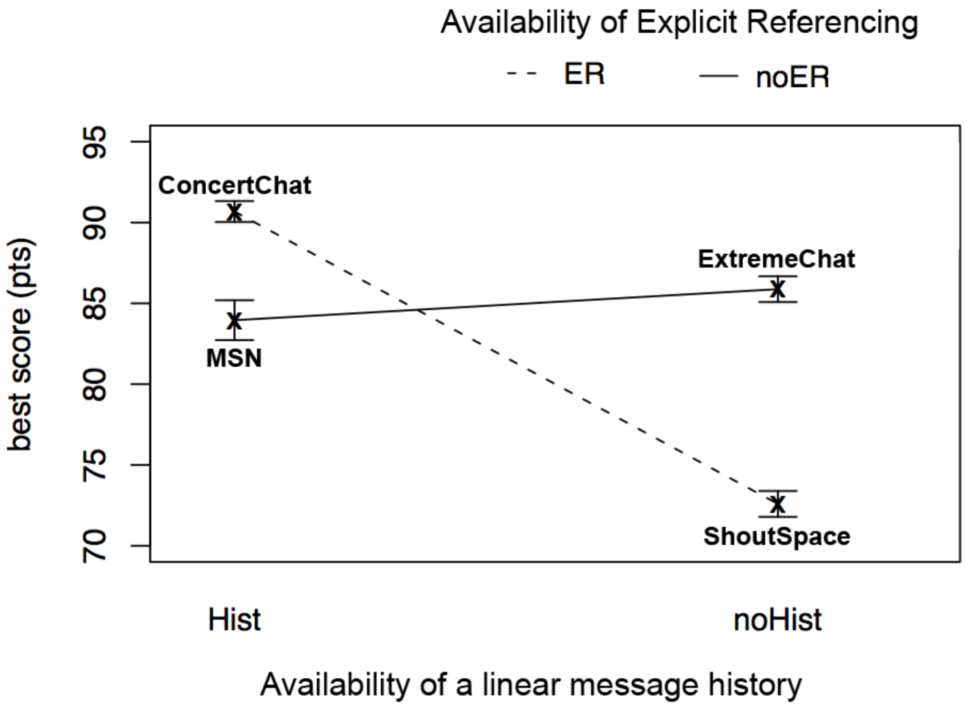


Figure 7-6: Interaction plot of Explicit Referencing × Message History on the best score achieved by the pair. The fences represent the standard error of mean

tool offered also a persistence of the conversation messages (for the interaction F[1,58]=124.38, p<.001). A TukeyHSD post hoc test confirmed that all the conditions differ when taken two by two (95% family-wise confidence level, factor levels ordered, p<.001), with the exception of the combination ExtremeChat–MSN (p>0.1, ns). See figure 7-6 for the interaction plot.

Icon initial positioning time

For all experiments, I measured the moment at which a pair managed to position the necessary icons on the map and then submitted the first solution for evaluation. In the majority of cases, participants began the task choosing the parking lots necessary to satisfy the provisionary audience. Then, the pairs typically positioned the stages and finally the order of concerts was decided, and the solution tested.

The availability of an Explicit Referencing mechanism did not have an impact on the time required by the teams to complete the initial positioning of the icons (F[1,38]=0.30, p>0.1, ns). This finding was not consistent with H1, as pairs using a tool supporting ER did not reach the initial positioning of the icons faster that pairs in the other conditions. However, this result was obtained averaging the scores obtained by ConcertChat (with the fastest positioning), with those of ShoutSpace (the slowest positioning).

The availability of a linear chat history had a large impact on the time required by the pairs for completing their initial positioning of the icons and the launching the evaluation. Pairs were faster in the trials in which there was a chat application with a linear message history, compared to those in which a linear history was not available (*Hist*: m= 837100, sd=615636 msec *vs. noHist*: m=1877000, sd=582956 msec; F[1,38]=62.44, p<.001). These findings were consistent with H2.

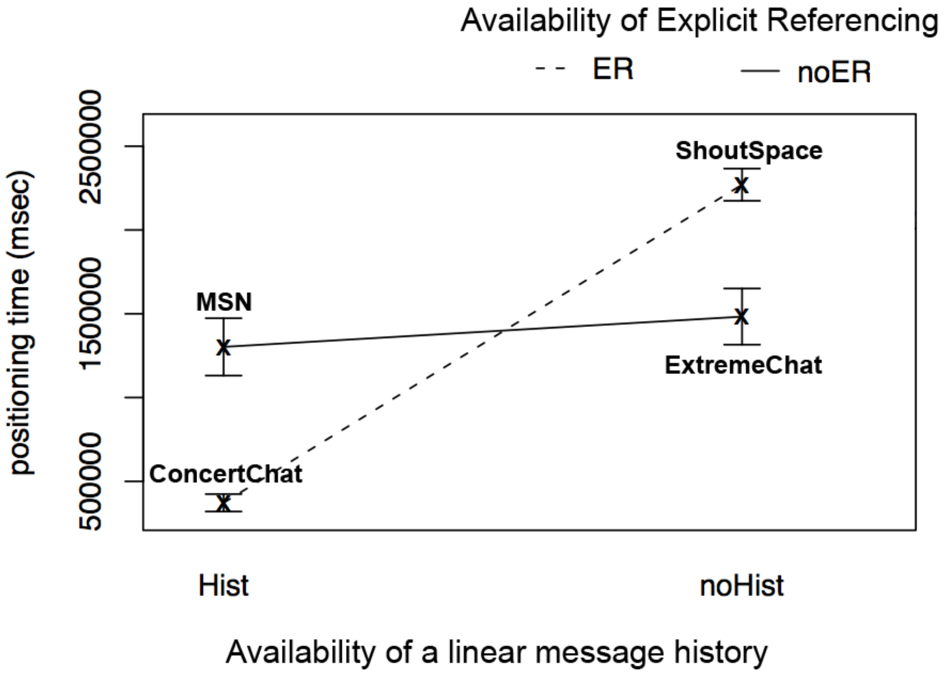


Figure 7-7: Interaction plot of Explicit Referencing × Message History on the icon initial positioning time. The fences represent the standard error of mean

The Explicit Referencing × Message History interaction showed that the impact of a Linear Message History on the time required to complete the initial positioning of the icons on the map was larger when an Explicit Referencing mechanism was present (for the interaction F[1,38]= 42.61, p<.001). A TukeyHSD post hoc test confirmed that all the conditions differ when taken two by two (95% family-wise confidence level, factor levels ordered, p<.001), with the exception of the combination ExtremeChat–MSN (p>0.1, ns). See figure 7-7 for the interaction plot.

Exploration space

As a final measure of performance, I considered how much of the solution space was explored by the pair. The positions that the icons could occupy on the map were a discrete number as were the possible ordering of concerts on the three stages. Each time that a pair evaluated a solution, I recorded three lists of the zones in which the icons were positioned: a list for the parking-icons, one for the stage-icons and the last one containing the sequence of stages for the concert-icons. For each experiment, I computed the sum of the substitutions in the lists of parking and stage icons. I added to this measure the number of permutations in the list of concert icons.

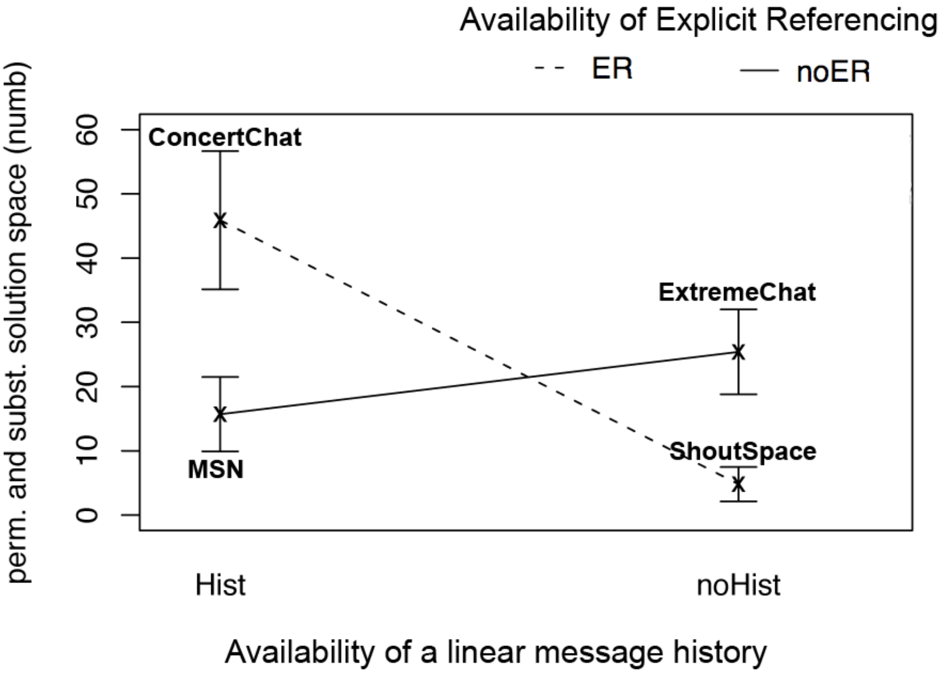


Figure 7-8: Interaction plot of Explicit Referencing × Message History on the exploration of the solution space. The fences represent the standard error of mean

The availability of an Explicit Referencing mechanism did not have an impact on the number of different zones used in all the solutions tested by a pair during an experiment (F[1,58]=0.41, p>0.1, ns). This result was not consistent with H1, but again this figure coalesced the statistics of participants using ConcertChat, who tested the highest percent of the solution space, with those of participants using ShoutSpace, who tested the smallest percent of the solution space.

The availability of a linear chat history had a large impact on the number of different zones used in all the solutions tested by a pair during an experiment. Pairs tested an higher percent of the solution space in the trials in which there was a chat application with a linear message history, compared to those in which a linear history was not available (*Hist*: m=28.07, sd=26.80 distinct zones *vs. noHist*: m=12.40, sd=17.48 distinct zones; F[1,58]=9.15, p<.005). These findings were consistent with H2.

The Explicit Referencing × Message History interaction showed that the impact of a linear message history on the number of different zones used in all the solutions tested by a pair during an experiment was larger when an Explicit Referencing mechanism was present (for the interaction F[1,58]=17.40, p<.001). A TukeyHSD post hoc test (95% family-wise confidence level, factor levels ordered, p<.001) revealed that the effect was influenced by the pair ConcertChat–ShoutSpace and ConcertChat–MSN. All the other combinations were not significant (p>.05, ns). See figure 7-8 for the interaction plot.

7.3.2 Conversational style

The third hypothesis that I defined previously postulates that: *H3, Explicit referencing makes communication more efficient (fewer sentences, with fewer words).* This section presents evidences to verify this hypothesis.

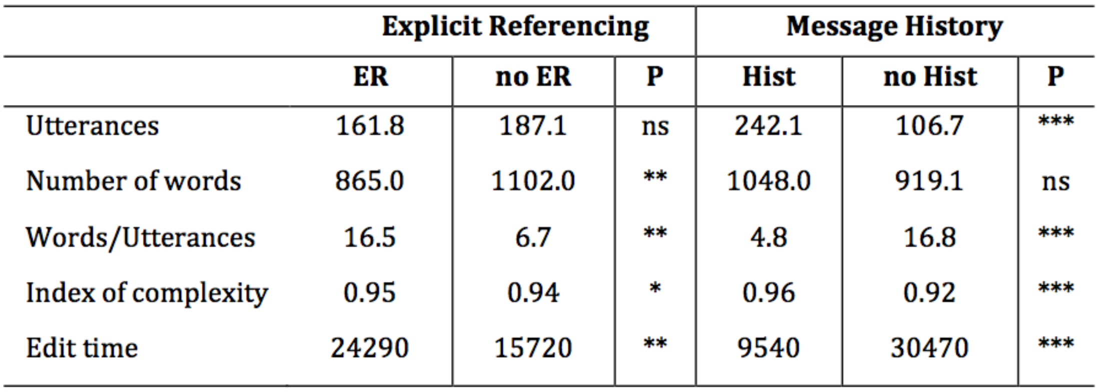
To understand how the pairs interacted in the different conditions, I computed a number of process variables focusing on the linguistic strategies employed by pairs to coordinate their efforts. I measured the quantity of speech produced by pairs and the regularity of turn-taking during the task. I also measured the amount of time needed to produce each utterance and the specific linguistic devices used to convey positioning meaning and manage the interaction. Finally, I computed specific markers to understand how participant pairs structured their interaction. In the rest of the chapter, I will use the word utterance to mean a single message sent through the chat application.

To understand how the media changed the structure of the conversation and its efficiency, I measured the number of utterances and the number of words in each experiment and the number of words per utterance for each condition. Additionally, to understand whether there was any asymmetry in the communication, I computed an index of complexity of turn taking (in short Ict)[3]: its value is 0 if, knowing the speaker at turn `n` I have a probability of 1 for predicting who will speak at `n+1`. Its value is 1 if knowing the speaker at turn `n` does not give us any information regarding who will speak at `n+1` (Dillenbourg et al., 1997; Lemay, 1999). Finally, I computed the time required by the participants to compose each message. The resulting data is summarized in table 7.3.

I chose to report only these measures because they correlate significantly with the score: the number of utterances is positively correlated with the pair score (r = 0.53, t[58]= 4.82, p<.001); the number of words per utterance is negatively correlated with the score (r = -0.66, t[58]= -6.75, p<.001). Mean editing time is negatively correlated with the score (r = -0.70, t[58]=-6.09, p<.001).

[3] The formula of this marker is described by Lemay (1999). See `http://tecfa.unige.ch/~lemay/thesis/`, last retrieved April 2008.

Table 7.3: Communication efficiency (***p<.001, **p<.01, *p<.05)



	Explicit Referencing			Message History		
	ER	no ER	P	Hist	no Hist	P
Utterances	161.8	187.1	ns	242.1	106.7	***
Number of words	865.0	1102.0	**	1048.0	919.1	ns
Words/Utterances	16.5	6.7	**	4.8	16.8	***
Index of complexity	0.95	0.94	*	0.96	0.92	***
Edit time	24290	15720	**	9540	30470	***

Effects of Explicit Referencing on efficiency and structure

The availability of the Explicit Referencing mechanism did not produce a significant difference in the number of utterances produced. The number of words was smaller when pairs could use ER compared to when they could not (*ER*: m=865.00, sd=323.14 *vs. noER*: m=1102.00, sd=323.43 words per experiment; F[1,58]=8.34, p<.006). This also reflected on the number of words per utterance that was significantly higher in the ER condition compared to the noER condition (*ER*: m=16.54, sd=33.47 *vs. noER*: m=6.65, sd=4.62 words per utterance; F[1,58]=9.75, p<.003). Finally, the index of complexity (Ict) shows that communication was more regular when participants could use an explicit referencing mechanism compared to the opposite condition (*ER*: m=0.95, sd=0.03 *vs. noER*: m=0.94, sd=0.04; F[1,58]=4.22, p<.05). These results are partially consistent with H2, which predicted more efficient communication with applications supporting ER. These findings are summarized in table 7.3.

These results are partially consistent with H3. Indeed, the number of utterances was smaller for those trials where participants could use a tool implementing ER. However, their number of words per utterance was higher. These results show that participants adopted different interaction styles, adapting their communication to the different constraints to which they were exposed.

Effects of linear message history on efficiency and structure

Pairs collaborating with a communication tool supporting a linear message history produced significantly more utterances during the experiment. These pairs doubled the number of utterances emitted (*Hist*: m=242.20, sd= 96.71 *vs. noHist*: m=103.70, sd=57.69 utterances per experiment; F[1,58]=53.00, p<.001). While this led to more words per experiment, the difference was not significant. On the contrary, I found the difference of words per utterance to be significant. Pairs with a linear message history employed almost a fourth of the words per message than pairs without a linear message history (*Hist*: m=4.81, sd=2.69 *vs.* *noHist*: m=16.84, sd=28.36 words per utterance; F[1,58]=81.86, p<.001). Finally, pairs with no linear message history had the least symmetrical turn taking compared to those with a linear message history (*Hist*: m=0.96, sd=0.02 *vs.* *noHist*: 0.92, sd=0.04; F[1,58]=33.46, p<.001).

These results suggest that participants adopted two dialogue styles: lots of short interwoven utterances for trials using **ConcertChat** and **MSN** and few, symmetric, and long contributions for trials using **ShoutSpace** and **ExtremeChat**. Therefore, H3 is not verified: the manipulation of the availability of Explicit Referencing did not result systematically in fewer utterances containing fewer words. This was the case for participants who communicated with ConcertChat, but it was not the case for participants who used ShoutSpace to solve the task.

Mean editing time of messages

Editing time of the messages was minimal when participants pairs used a chat tool supporting Explicit Referencing and with a linear message history. When I varied the availability of a linear message history, I observed that the mean editing time nearly tripled in length (*Hist*: m=9540, sd=4832 *vs.* *noHist*: m=30470, sd=16650 msec; F[1,38]=52.64, p<.001). When I varied the availability of the Explicit Referencing mechanism, I observed that participants without ER were faster in composing their messages than those with ER (*ER*: m=24290, sd=20438 *vs.* *noER*: m=15720, sd=8644 msec; F[1,38]= 8.82, p<.006).

The Explicit Referencing × Message History interaction showed that the impact of a Linear Message History on the mean time required to compose the messages used in the experiment was longer when an Explicit Referencing mechanism was present (for the interaction F[1,38]=23.84, p<.001). A TukeyHSD post hoc test confirmed that all the conditions differ when taken two by two (95% family-wise confidence level, factor levels ordered, p<.001), with the exception of the combination ConcertChat–MSN (p=0.54, ns) and the combination ExtremeChat–MSN (p=0.35, ns). See figure 7-9 for the interaction plot.

Again, H3 is not verified by these results. While the hypothesis predicted correctly a little emission effort for messages produced using ConcertChat, it did not predict greater effort for messages produced with ShoutSpace.

7.3.3 Linguistic spatial positioning

As the task required precise positioning on the shared map, I analyzed the language employed by the participants. I processed each message using an automatic method and a manual categorization (presented in section 7.2.4). The automatic feature extraction was operated using TreeTagger (Schmid, 1994). The algorithm handled the stemming of each word and the tagging according to

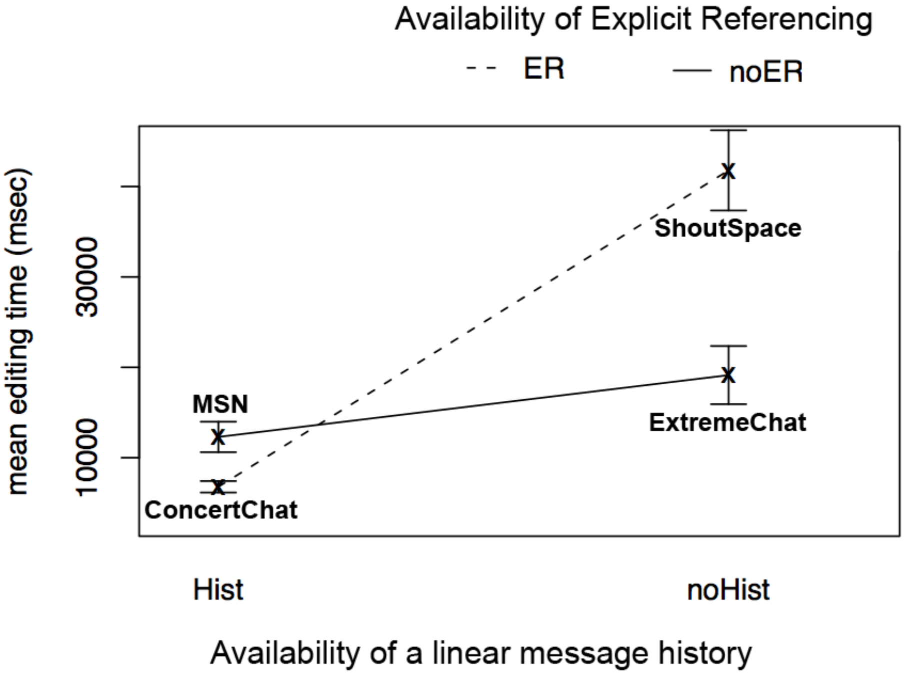


Figure 7-9: Interaction plot of Explicit Referencing × Message History on the mean editing time of messages. The fences represent the standard error of mean

the French parameter file[4].

Automatic Features Extractions From the Corpus

I used three features of the participants' language to measure how different linguistic devices were used in function with the different media. I counted the number of prepositional phrases (e.g., "*on the right hand-side of the parking lot*", "*below the 'H'-shaped building*", etc.) that were employed by the participants in their conversation. Similarly, I counted the number of spatial adverbial clauses that were used as positioning device in the dialogue (e.g., "*leave the parking icon where it is*"). Finally, I counted the number of deictic expressions used in the conversation (e.g., "*I placed the second concert here*", or "*move your icon there*"). Table 7.4 presents the resulting data.

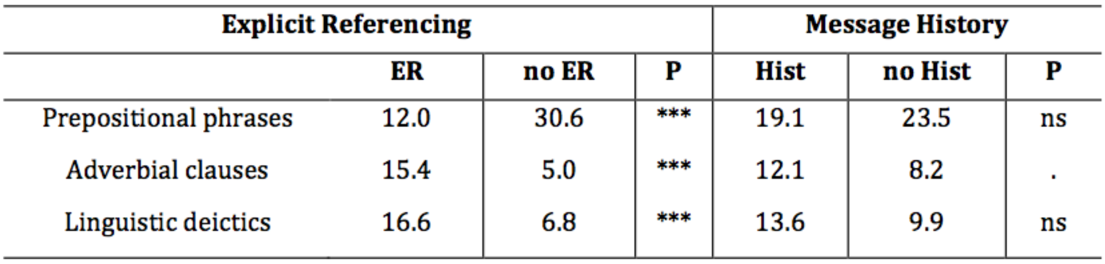
Pairs solving the task with a communication tool implementing a linear message history did not produce any significant difference in these three categories. On the contrary, when I varied the availability of the Explicit Referencing mechanism, I observed significant differences across the three linguistic features. Participant pairs using tools that implemented an Explicit Referencing mechanism used almost three times fewer of the prepositional phrases than the

[4] The documentation of these components, as well as the source code of the tagger is available at `http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/`, last retrieved February 2008.

pairs using tools without ER (*ER*: m=12.03, sd=8.21 *vs.* *noER*: m=30.57, sd=13.04 prepositional phrases per experiment; F[1,58]=43.84, p<.001). On the other hand, participants with ER produced three times more place adverbial clauses than participants with no ER mechanism available (*ER*: m=15.37, sd=11.13 *vs.* *noER*: m=4.37, sd=4.34 adverbial clauses per experiment; F[1,58]=23.54, p<.001). Lastly, participants with ER produced two times more linguistic deictic expressions than participants with no Explicit Referencing tool (*ER*: m=16.60, sd=11.99 *vs.* *noER*: m=6.83, sd=3.26 deictic expressions per experiment; F[1,58]=19.07, p<.001). These values are visualized in figure 7-10.

Consistent with the findings reported in the previous section, these results suggests that participants adapt their communication strategy to reduce their grounding effort. Participants using tools implementing ER adapted accordingly their communication reducing prepositional phrases which take more effort to encode and which are more prone to generate mistakes and miscomprehension.

Table 7.4: Linguistic Spatial Positioning (***p<.001, .p<0.1)



	Explicit Referencing			Message History		
	ER	no ER	P	Hist	no Hist	P
Prepositional phrases	12.0	30.6	***	19.1	23.5	ns
Adverbial clauses	15.4	5.0	***	12.1	8.2	.
Linguistic deictics	16.6	6.8	***	13.6	9.9	ns

Interaction coding system

As explained in section 7.2.4, a unique and mutually exclusive category code was associated to each corpus segment and was systematically counted for each participant (two researchers analyzed the interactions; inter-coder reliability was good: Kappa = .81, p < .0001).

The aim of the coding scheme was to highlight possible differences between the strategies that participants had to adopt to solve the task in the different experimental conditions. However, the manipulation of the experimental conditions did not produce any statistical difference in the content categories used to classify the messages produced in the experiments.

As a second goal, the coding scheme classified the different strategies used by participants to position the icon over the map. To eliminate the influence of the uneven number of messages across conditions, I divided each tag frequency reported here by the total number of messages in the experiment. Pairs solving the task with a communication tool supporting a linear message history did not produce any significant difference in these four categories. On the contrary, pairs

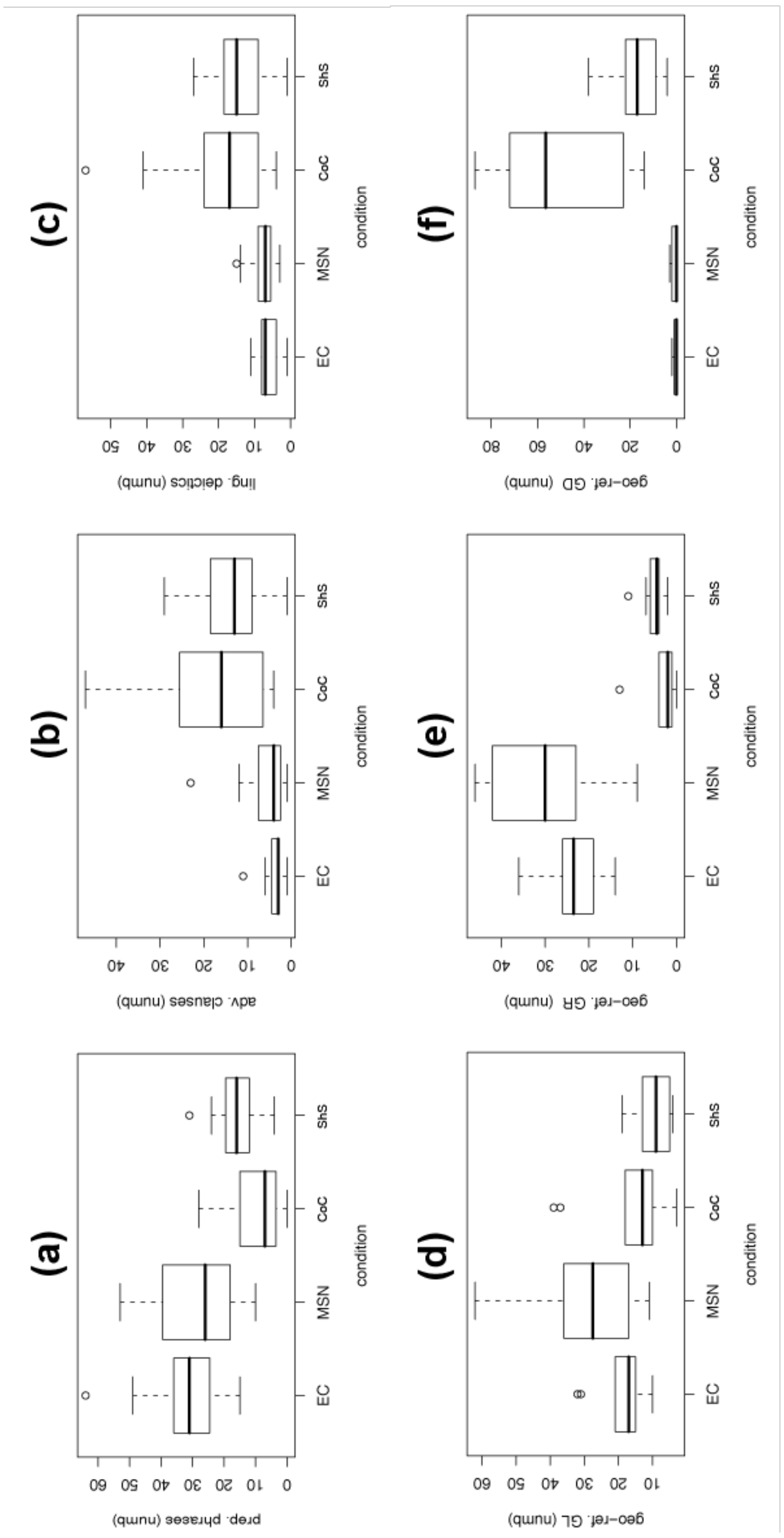


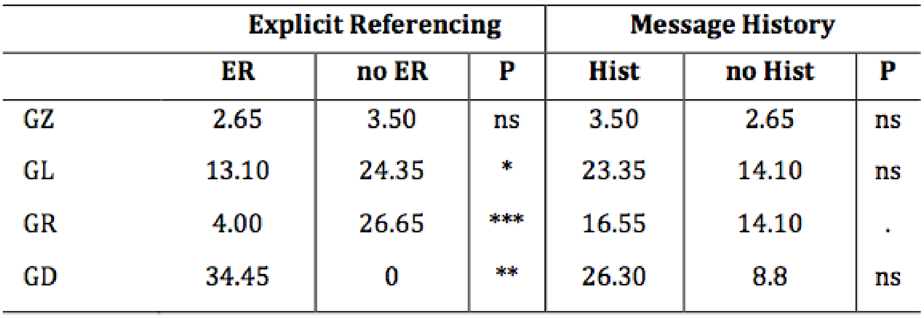
Figure 7-10: Boxplots of the three linguistic features automatically extracted from the corpus (right) and strategies for placing the icons tagged manually using the scheme presented in section 7.2.4 (right). (a) Prepositional phrases; (b) adverbial clauses; (c) linguistic deictics; (d) georeference of the kind GL; (e) kind GR; and (f) kind GD

solving the task with a communication tool implementing the Explicit Referencing mechanism produced significant differences across three of the four categories. The manipulation of the factors did not produce any difference in the amount of positioning references of the kind GZ (p>0.9, ns).

Participant pairs using tools that allowed for Explicit Referencing used fewer phrases tagged as GL than the pairs without ER (*ER*: m=13.10, sd=9.67 *vs.* *noER*: m=24.35, sd=12.87 messages tagged GL/total messages; F[1,38]= 7.28, p<.01). Participants with ER produced six times fewer messages tagged as GR than participants with no ER mechanism available (*ER*: m=4.00, sd=11.13 *vs.* *noER*: m=26.65, sd=4.34 messages tagged GR/total messages; F[1,38]=68.72, p<.001). It is important to notice that for this particular category, the manipulation of the availability of a linear message history had an impact that could almost be considered significant: participants without a linear message history produced fewer messages tagged as GR than participants in the complementary condition (*Hist*: m=16.55, sd=17.03 *vs.* *noHist*: m=14.10, sd=10.44 messages tagged GR/total messages; F[1,38]= 3.37, p=0.07, ns). These values are visualized in figure 7-10.

Lastly, participants with ER employed actively Explicit Referencing as an alternative to other positioning strategies (*ER*: m=34.45, sd=26.41 *vs.* *noER*: m=0, sd=0 messages tagged GD/total messages; F[1,38]= 8.13, p<.001). Table 7.5 summarizes these results.

Table 7.5: Spatial Positioning Strategies (***p<.001, **p<.01, *p<.05, .p<0.1)



	Explicit Referencing			Message History		
	ER	no ER	P	Hist	no Hist	P
GZ	2.65	3.50	ns	3.50	2.65	ns
GL	13.10	24.35	*	23.35	14.10	ns
GR	4.00	26.65	***	16.55	14.10	.
GD	34.45	0	**	26.30	8.8	ns

These results demonstrate that participants appropriated the communication tool that they used to collaborate. Participants who could use deictic references over the shared map preferred this modality to using longer and more complex descriptions (e.g., geo-references of the kind GR). Additionally, these results show that the availability of ER did not modify the way participant managed their interaction to solve the 'festival' task.

7.3.4 Icon position mistakes

As an indication of the quality of collaboration, I measured the number of mistakes that participants produced in placing the icons in different zones of the map. The positioning differences were computed each time a score evaluation was invoked by one of the participants. A high number of mistakes were caused by participants independently testing solutions. This behavior was clearly forbidden in the assignments, but I nevertheless found 13 experiments, over a total of 40 experiments that I tagged manually, where this occurred.

Participant pairs using tools that allowed for Explicit Referencing produced almost a third of the icon position mistakes than those of the pairs without ER (*ER*: m=12.00, sd=20.86 *vs. noER*: m=33.60, sd=22.79 mistakes; F[1,38]=10.36, p<.01). It is important to notice that the manipulation of the availability of a linear message history had an impact that approached being significant: participants with a linear message history produced fewer icon position mistakes than participants in the complementary condition (*Hist*: m=16.10, sd=21.24 *vs. noHist*: m=29.50, sd=25.58 mistakes; F[1,38]=3.98, p= 0.053, ns).

7.3.5 Mediation effects

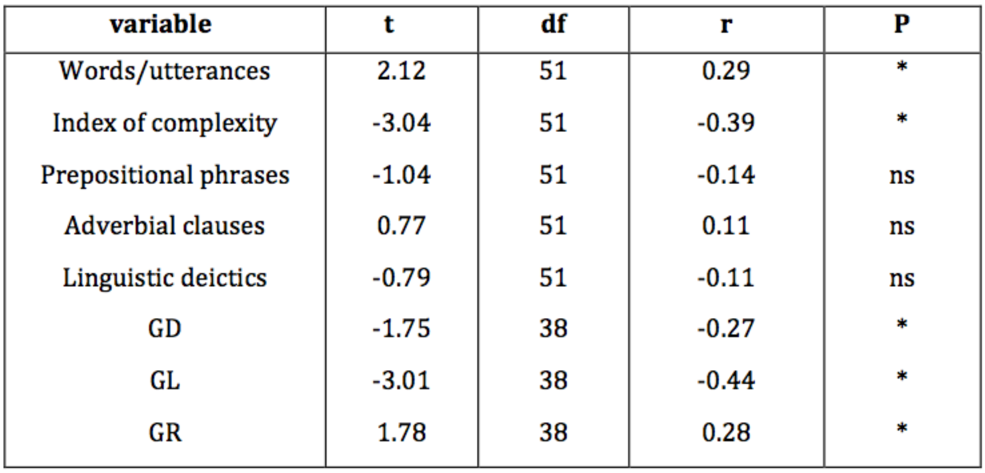
Interaction variables were examined as variables potentially mediating the effect of the experimental condition on the best scores achieved by the participant pairs. Regression analyses were performed to assess mediating effects using regression methods after the methodology developed by Baron and Kenny (1986).

The regression test confirmed that the message mean editing time variable was significantly and negatively related to the availability of a linear message history (β =-0.624, p<.001). Additionally, the mean editing time was significantly and positively related to the availability of an Explicit Referencing mechanism (β =.310, p<.01) and significantly and negatively related to the best score (β =-0.655, p<.001). These effects disappeared when I tested the prediction of the best score with the linear model combining the factors and the mean editing time. While the mediation variable was still significant (β =-0.499, p<.01), the factors were not (*Hist*: β =.190, p=0.19, ns; ER: β =-0.09, p=0.40, ns). Finally I tested the significance of the found mediation effect using the test developed by Sobel (1982). The test confirmed the effect for both factors (*Hist*: t=-6,25, p<.001; ER: t=3.11, p<.01).

The mean editing time appears to be a valid mediation variable that can account for the effects caused by the experimental manipulation of the independent variable on the results recorded for the dependent variables. To better understand how the editing time of a message was related to the linguistic characteristics of the message itself, I tested the correlation between the mean editing time and the other process variables related to the linguistic qualities of the message that had an effect on the dependent variables. For the test reported subsequently, I used the correlation test of Pearson (Best & Roberts, 1975). The number of words per utterance was positively correlated with the mean editing time (r = 0.29, t[51]= 2.13, p<.05). The index of complexity defined above was negatively correlated with the mean editing time (r = -0.39, t[51]=

-3.04, p<.05). The number of prepositional phrases, as well as the number of adverbial clauses and the number of deictic expressions were not correlated with the mean editing time (the actual values of the test are reported in table 7.6). The number of 'deictic' geo-references in each experiment (GD) was negatively correlated with the mean editing time (r = -0.27, t[38]= -1.75, p<.05). The number of 'label' geo-references (GL) was also negatively correlated with the mean editing time (r = -0.44, t[38]= -3.01, p<.05). Finally, the number of 'relative' geo-references was positively correlated to the mean editing time (r =0.28, t[38]= 1.78, p<.05). Table 7.6 summarizes these results.

Table 7.6: Pearson's product-moment correlations of the linguistic variables with the mean editing time (*p<.05)



variable	t	df	r	P
Words/utterances	2.12	51	0.29	*
Index of complexity	-3.04	51	-0.39	*
Prepositional phrases	-1.04	51	-0.14	ns
Adverbial clauses	0.77	51	0.11	ns
Linguistic deictics	-0.79	51	-0.11	ns
GD	-1.75	38	-0.27	*
GL	-3.01	38	-0.44	*
GR	1.78	38	0.28	*

This findings are not consistent with H3, which was predicting that pairs solving the task with collaboration tools implementing ER would communicate more efficiently. The mean editing time mediates the effects of the other process variables and is positively related to the availability of ER. Pairs using tools implementing ER took more time to edit the messages. However, this results is the outcome of the average of the mean editing times obtained by pairs using ShoutSpace (the longest) with those obtained by pairs using ConcertChat (the shortest).

7.3.6 Qualitative descriptions

In this section, I will present excerpts from conversations occurred during the experiments to demonstrate qualitatively how participants adapted to the different experimental conditions. For each text, I will indicate the experimental condition of the experiment. The transcripts reported below were translated from French. I also tried to replicate original mistakes, typos, and other linguistic subtleties that were employed by the participants to give a sense of the miscomprehension that reading these messages could generate. Why could participant pairs complete their experiment with fewer utterances when they did not have a message history? Without message permanence, participants were more systematic in taking the necessary steps to solve the task. They used more words per message, and employed more regular turn taking in order to minimize information loss.

Excerpt 1.1 reports two fragments of the conversation between the pairs when they were using ShoutSpace or MSN chat. The pairs were, in both cases, trying to position a stage on the plan. The interaction lasted about the same time in both cases. This excerpt demonstrates how the pair took advantage of the spatial position at which the messages were positioned in ShoutSpace to reduce the information conveyed by the text. It also demonstrates how the presence of the history brought the pair to alternate different subjects of discussion.

EXCERPT 1.1, experiment 13, **ShouSpace** condition.

```
[1] A: Show me where with the messages and I will put the things at the same
    place
[2] A: Show me where you put the second stage
[3] [B post an empty message on the map]
[4] B: Look on the top-left corner, I posted a message
[5] A: 2nd Stage. Ok, I set the 3 parkings, the central stage and I am going
    to set this one
[6] B: 2nd Stage. Where do you want to put it exactly?
[7] A: Re: 2nd Stage. Ok. and for the third?
```

EXCERPT 1.1 –continued, experiment 12, **MSN** condition.

```
[1] B: we can place the stages on the right hand side
[2] A: Ok
[3] A: Let's put one just on the right side
[4] A: close to the dashed line?
[5] B: I am not following you
[6] A: do you see the small black arrow? On the right?
[7] B: Yes
[8] A: Just a bit over that
[9] B: Ok
[10] B: On the right side of the dashed line
[11] A: before the little route deviation
[12] A: yes
[13] B: Ok
```

```
[14] B: for the parkings, we need 900 spots. So, we take all the lots close
    to this area
```

The tools supporting a linear message history allowed users to have a more 'flexible' turn-taking. Sentences started on a certain message could be interrupted and sent incomplete to the other participant and then completed by a second message (see for instance messages 6-7, or 9-10 in excerpt 1.2). Of course, with this kind of flexibility, it frequently happened that multiple intentions were intertwined in the message exchange and participants felt lost. The co-text loss in chat conversations refers to situations in which a participant does not establish a conversation thread (Pimentel et al., 2003, p. 484): *"co-text loss occurs each time the reader is unable to identify which of the previous messages provides the elements that are necessary to understand the message that is being read"*. In the **MSN** condition, this happened quite frequently, as it can be seen in the excerpt 1.2 at line 10-17. Here participant B is proposing a new location for a scene. A is uncertain about what the other is talking about and asks a clarification (line 13). The other does not realize immediately that the collaborator is lost. So she goes on with sentence 15. Finally she provide a clarification on her intentions (line 17).

EXCERPT 1.2, experiment 20, **MSN** condition.

```
[1] A: there is a path going up to the first scene on the right of this path
    between the two crossing
[2] A: at the same height than P450
[3] B: 2 sec
[4] B: but this place is small
[5] A: yes if this works good otherwise we can put ir on the place that is
    to the left of the place we are taking about
[6] B: we do not have to put it in the hole which is on the left
[7] B: but kind 3cm on the south?
[8] A: no
[9] B: OK and what about kind
[10] B: between the 3 crossings all the way on the left
[11] B: (I did not understood the crossings on the bottom of the screen)
[12] B: OK?
[13] A: No, I do not understand? Are you proposing a new place for the
    last stage?
[14] B: Yes? You are not?
[15] B: we should maybe have only 2 and shit
[16] B: ?
```

[17] A: yes yes: I propose on the left of the trail (which is all the way on the left of the map) on the same height of P450

[18] A: an you what do you propose

[19] B: in the hole?

[20] A: hole?

[21] B: ok, on the left of p200 there is a 'hole'

[22] B: yes?

Confusion was not only typical of the **MSN** condition but also of the **ConcertChat** condition. Even if participants had at their disposal the Explicit Referencing mechanism, this did not help prevent confusion generated by intertwined turns. Excerpt 1.3 provides an example of such situation. Even if participant A had visual access to the list of proposed parkings proposed by participant B, she asked again the collaborator to repeat them (line 10) as she felt unsure of their locations after the exchange that had occurred a few messages before.

EXCERPT 1.3, experiment 43, **ConcertChat** condition.

[1] A: after we could take this one and we can fforget the parkings for now [Msg has a reference to the map]

[2] A: are you ok?

[3] B: yeas but it is far.s

[4] B: we can take the 4 parkings at the top

[5] B: 210 + 190 + 175 + 435

[6] B: 100+100+350+750 chf

[7] A: on p350??

[8] B: this gives the 900 places for 1300chf

[9] A: this is cool!!

[10] A: but say again which ones

Of course, ER was actively used by the pairs using tools supporting it. Excerpt 1.4 shows how participants who had at their disposal an Explicit Referencing tool preferred using deictic acts instead of more elaborated linguistic descriptions that could potentially lead to miscomprehensions. This fragment also shows how participants used the only labels present on the map, namely those showing parking lots capacity, as a grounding anchor to situate their discussion (line 4, and 6). For the stages they usually used a custom-made code, generally using the label indicating the capacity of the closest parking lot.

EXCERPT 1.4, experiment 45, **ConcertChat** condition.

```
[1] A: first of all the stages
[2] A: Where do you want to position them?
[3] B: I' d say one between the parking 195 and 120
[4] B: one below parking 450
[5] A: here? [Msg has a reference to the map]
[6] B: and the last on the side of park 30
[7] B: here [Msg has a reference to the map]
[8] B: here [Msg has a reference to the map]
[9] B: here [Msg has a reference to the map]
[10] A: here [Msg has a reference to the map]
[11] A: OK!
[12] A: Ok for me
```

Participants in the **ShoutSpace** condition had at their disposal the ER mechanism. However, participant had a hard time to acknowledging the messages of their collaborators and to following the natural flow of the conversation. To overcome these limitations, many participants in this condition put in place strategies to help them cope with the communication difficulties they had with the tool. Excerpt 1.5 demonstrate this point. At the beginning of the conversation, participants posted messages on locations (line 1-2) but they they switched to using labels as it was more practical to indicate multiple locations (line 3, and 7). Also, it is possible to see how participants were actively looking for acknowledgement that their partner had understood the message as they were unsure on whether their collaborator had red one of their previous contributions (line 6). Figure 7-11 shows a screen capture of a portion of one of the participant's display of the experiment 17. The area presents some of the messages contained in the excerpt 1.5 that have been 'despatialized' as they did not refer to specific locations of the map. Participant in the **ShoutSpace** condition often placed these messages to the side of the campus map, over regions that were not used to position the icons.

EXCERPT 1.5, experiment 17, **ShoutSpace** condition.

```
[All msgs have a reference to the map]
[1] A: Re: parking: I propose this parking!
[2] A: And also this one
[3] B: Re: parking:ok, also this one and that for 175 people and that for 180
    people on the corner
[4] B: Re: ahm...ok!
```

[5] A: 3rd parking: This one has a good ration between the capacity and the price!

[6] B: Re:3rd parking: Yes. Did you read my answer on the parking for 185 people?

[7] A: Re: parking recap!: So, let's recap: 435 places, 750 CHF 175places, 350 CHF 180 places, 200 CHF For the last one, there is also that on the north west with 190 places and 100CHF, not expensive and not far!!!

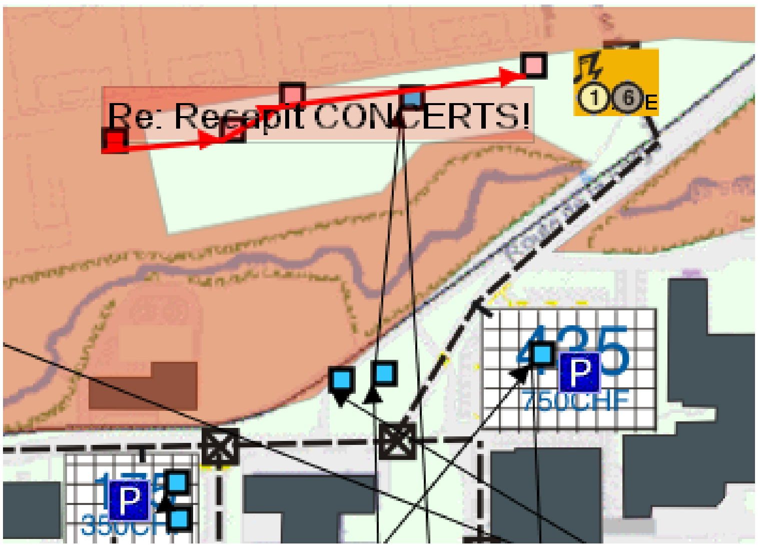


Figure 7-11: Screen capture of the experiment 17, condition **ShoutSpace**. The area shows some of the messages presented in the excerpt 1.5 that have been *despatialized* by the participants as they did not refer to a specific location

Excerpt 1.6 exemplifies how the lack of the message history influenced the way participants had to structure their contributions in order to be understood: each message had to be self-explanatory as previous utterances might be not relevant any more for the message recipient (line 2). Participants in the **ExtremeChat** condition had to find strategies to overcome the limitations of the tool they were using to collaborate. Excerpt 1.7 shows how participants in this condition names scenes with letters to facilitate coordination of movements of the icons. Participant used A, B, and C to name the three stages. They they could discuss the order of the concerts simply saying: "the first in C" and so forth. This excerpt shows also the regularities of the turn-taking in the **ExtremeChat** condition.

EXCERPT 1.6, experiment 57, **ExtremeChat** condition.

[1] A: did you position your stage?

```
[2] B: my stage is on the squared space that is formed by the access routes
    on the right. Where is yours?
[3] A: OK. For my stage... you take the big parking 435. you take its lower-
    left corner and you paste the upper-right corner of the stage icon to this
    one. Do you see what I mean?
[4] A: there is an H-shaped building. My stage is on its right hand side
[5] A: is that ok?
[6] B: yours is fine but now mine is too close to yours
[7] A: ok
```

EXCERPT 1.7, experiment 59, **ExtremeChat** condition.

```
[1] A: so first concert stage B ok?
[2] B: yes, the second in c
[3] A: yes, 3 B
[4] B: the third in B
[5] B: 4 in A
[6] A: 5 in B
[7] B: exact
[8] A: and 6 in c
[9] B: 6 in C
[10] A: ok
[11] B: score?
```

Of course, besides these simple strategies, participants in the **ExtremeChat** condition had to employ complex descriptions of the places in order to position the stages. Excerpt 1.8 presents a sequence of messages coded as Position Providing (PP) and Position Seeking (PS) that the two participants alternated to reach the correct placement of a stage. In the example, participant A chose to use the shape of a building that resemble the letter 'H' (line 1). However, the strategy did not work well as the collaborator thought about an empty spot to the left of this building (line 5). So to reach the correct position, participants had to go through a series of approximations.

It is also interesting to notice how participant B used a copy/paste command at line 2 to quote the previous message, as she knew that her partner would not have access to the history.

EXCERPT 1.8, experiment 35, **ExtremeChat** condition.

```
[1] A: (for the 2nd park: close to the left of the cross) for the 3rd park
    between the two intersection on the left before the building in the
    form of an H
```

```
[2] B: ‘‘for the 3rd park between the two intersection on the left before
    the building in the form of an H’’ so between the park 175 and 185?
[3] A: yes, exact
[4] A: ok?
[5] B: so at the median intersection between 175 and 185, below on
   the left of H. I am placing the icon
[6] A: ah NO!
[7] B: bah so where
[8] B: u talk of the strange H
[9] A: sorry, my mistake. below park 185 just on the other side of the
    street this is what I ment
[10] A: yes
```

Finally, participants that collaborated with communication tools that did not support Explicit Referencing but that had a linear message history could also use referencing to the previous messages taking advantage of the persistence of the previous utterances (a mechanism that is defined as anaphora, as explained in chapter 2). This behavior was not frequent overall, however excerpt 1.9 shows an instance of such attitude. Participant `B` proposed a list of parking lots to the collaborator. This collaborator, after a few messages agreed to the original idea and instead of re-typing the list made a reference to the original proposition (line 5).

EXCERPT 1.9, experiment 14, **MSN** condition.

```
[1] B: recap: we take on the right, 435, 185, 180 et 175, OK?
[2] A: ya
[3] A: And what do you think of placing a stage below park with 435 places?
[4] B: Do we place the icons or we wait?
[5] A: Let’s go with the parking icons, on the center, on those that you
    mentioned
```

7.4 Discussion

Q5, does the availability of Explicit Referencing enhance the performance in a collaborative problem solving task at a distance?

The results presented in this work demonstrate that in collaborative tasks at distance that involve problem solving on shared maps, Explicit Referencing improves teams performance. However if the implementation of this mechanism breaks the linearity of the conversation context (e.g., organising messages according to the space), performance is reduced.

H1, Explicit referencing leads to better team performance.
This hypothesis was verified if we consider separately the results obtained by pairs solving the task with ConcertChat, the best, from those using ShoutSpace, the worst. Pairs using ConcertChat obtained the best performance in the three measures described in this chapter. Results here presented suggest the idea that improvements of performance of ConcertChat might be due to two factors: (a) while the other tools give support only to the conversation-context or only the task-context, ConcertChat give support to both; (b) participants using ConcertChat have at their disposal two modalities of communication (with ER and without) while the other had only one (with ER or without).

(a) My initial idea was that participants in the **MSN** or the **ExtremeChat** condition would perform worse than those in the other conditions because of the difficulty they could experience in disambiguating verbal descriptions of map locations. The results of the 'festival' experiment refine this initial assumption. The two communication contexts introduced at the beginning of this chapter, the conversation-context and the task-context, influenced different aspects of the collaboration. Actually, some utterances (for instance, about positioning parking icons) do refer to map locations while others, namely discussing the task strategy or task management, are independent of any location. The interface of ShoutSpace, like other commercial analogue interfaces discussed in chapter 4, fosters links of the conversation to the task-context to the detriment of maintaining the coherence of conversation-context. MSN chat and ExtremeChat have the inverse effect, not supporting Explicit Referencing. ConcertChat, on the other hand, supports both sides of the interaction context.

(b) From a qualitative analysis of the experiments, I observed that participants preferred to express task-management utterances via classical chat and used the shared pointers for the action-related utterances, when available. This appears to be consistent with the results reported by Dillenbourg and Traum (2006), who compared the content exchanged on a whiteboard and in a chat application in a collaborative problem solving. They found that users spontaneously display the more persistent information on the more persistent medium (the whiteboard), which serves as a group working memory. Cross-referencing the verbal utterances and whiteboard objects was mainly achieved by explicit verbal references and subtle timing cues.

H2, CAS organizing messages according to the position on the map to which they refer lead to inferior performance.
This hypotesis was verified. Pairs using ShoutSpace to solve the task obtained systematically

the worst results. Even participants using ExtremeChat, which did not offer any ER support and in which messages did not follow a temporal order, obtained better results than participants using ShoutSpace. What emerges from the analysis reported in this chapter is that more than the permanence of the conversation, what impacted negatively over the interaction of participants using ShoutSpace was the difficulty they had in following the flow of the conversation. Given the spatial dispersion of the conversation, it was difficult for them to be aware of incoming messages.

Participants in the **ShoutSpace** condition were forced to 'de-spatialize' management utterances, moving these messages over non-functional parts of the map in such a way so as to not interfere with other important information shown on the map. Also, the fragmentation of the conversation resulted in fewer messages with a higher number of words per message. In other words, participants tried to reconstruct the conversation-context by increasing the words for each message and decreasing the number of messages, since moving to the next one required an extra effort for reconstructing, in the participant's mind, the conversation-context. While in the **MSN** condition, I often observed that multiple conversations were concatenated (sometimes with a resulting confusion), participants in the **ShS** condition often reported a difficulty in following multiple conversations at the same time. This led them to a more structured turn taking to prevent confusion.

Conversely, participants in the **MSN** or **ExtremeChat** condition had to develop positioning and routing strategies for placing the icons. When interviewed after the task, almost all of the participants in these conditions noticed that the capacities of the parking lots, marked on the map, were unique numbers and used these as anchors for disambiguating the positioning with the partner. In contrast, participants in the **ShoutSpace** condition rarely noticed these unique labels. As the placement of the stages was often done after that of the parking lots, these were used as landmarks and as routing points for the positioning of the stages. This resulted as an increased use of prepositional phrases or adverbial clauses. Participants in this condition not only took full advantage of the parking lot capacity labels, using these to easily indicate which parking lots to select to the partner but also to sub-divide the map space so as to reduce confusion (e.g., "*the stage close to P140*").

These results can be understood by considering the Distributed Cognition framework (E. L. Hutchins & Klausen, 1991). The participants, the map and the communication tool can be conceptualized as being part of the same cognitive grid. A cognitive system adapts dynamically to changes and restrictions. Systems with pairs using ConcertChat could offload events and information through the shared workspace, while those in systems with pairs using ExtremeChat were forced to slow the communication peace, by including more information in each message and striving for the maximum of clarity. These systems obtained a comparable efficiency to those of pairs using MSN. System of pairs using ShoutSpace, could not compensate enough for the restrictions imposed by the communication tool and therefore they obtained inferior results.

Finally, results here presented are not consistent with those presented by Gergle et al. (2004) on the effects of chat persistence of collaborative performance. While they reported an improvement of the performance for those pairs with message persistence in their experiment, I did not find the same spread between the condition **ExtremeChat** and **MSN**. In my analysis of the messages, I observed that participants adapted to the restriction imposed by the application through modifying their communication strategies (more words per messages, with salient facts, and with a slower pace). It is important to notice that I designed a task that was more complex than the task used by these authors. In this regard, the presented results seem more similar to those of McCarthy and Monk (1994).

H3, Explicit referencing makes communication more efficient (fewer sentences, with fewer words).
As in the case of H1, this hypothesis was verified if we consider separately the results obtained by pairs using ConcertChat and those obtained by pairs using ShoutSpace. As I will discuss below (section 7.4.1) participants using ConcertChat took the shortest amount of time to edit their messages and therefore we can roughly say that they put the smallest effort in the sentence production. However, the results here presented challenge the definition of communication efficiency that I employed. A smaller amount of sentences with a smaller amount of words is not enough to say that this communication is more efficient. Human communication is adaptive to restrained conditions, as discussed in chapter 2. Perhaps, more subtle definition of *communication effort* is necessary.

7.4.1 Interpretation of the results

The festival experiment confronted the participants with a number of constraints that needed to be optimized. The optimization of these constraints was artificially made difficult by opposing the criteria needed to fulfill the task. Additionally the optimization of these contraints required the testing of multiple solutions to learn the features of the game that had an influence on these parameters. For instance, the presence of buildings between two stages did not reduce the noise of the sound checks. This fact, among many others had to be learned during the game, as they were explicitly not communicated during the training. Submitting multiple solutions was not only encouraged during the training, but also required in order for the participants to understand the features that could influence the evaluation of the proposed solution. Finding the best solution was essentially an approximation through trials and errors and therefore the more solutions submitted the higher the chance to find configurations that were rated high by the score tool.

Testing multiple solutions required time, as one of the most stressed requirements was that of minimizing icon position mistakes across the two maps. Each time the users wanted to check a certain configuration, the collaborators had to finely coordinate the movements of the icons using the communication tool assigned to their experimental condition. Some pairs used a systematic approach, like changing one factor at a time while keep all the rest constant. Others tried completely different configurations at each score, try to exploit extreme points of the solution space. Most of the pairs developed strategies to route the objects on the map. Most of the participants that could not point used the label of the parking lots as a readily available coordination device. Other participants relied on different strategies like that of using the names of the buildings (they relied on their common knowledge of the campus as these names were not available on the map used during the trials), or that of mentally subdividing the map in sectors and using these invisible coordinates to reduce to the zone of interest.

The communication tool used in the different conditions impacted directly on the time required to compose the messages. While in ConcertChat participants could just say "*here*" to show to their partner the new location of an icon, this was not possible for participants using MSN. These had to describe the place in different manners. Those who had a strategy could use it to reduce the time requited to compose a single message (e.g., "*cell a-5*", or "*North P80*"). On the contrary, those who did not employ shortcuts had to go through complex descriptions to synchronize the movements of the icons (e.g., "See the parking 250? There is a building on the right-hand side. Two centimeters below that building."). As explained in section 7.3.5, the message mean editing time appears to be a valid mediation variable that explains the effect of the manipulation of the experimental conditions on the dependent variables.

The other process variables that I have presented in the rest of the analysis are summarized by the editing time. None of them can singly explain the effect of the experimental manipulation on the obtained results. However, these variables contribute to explaining the message mean editing time. Quite obviously, the higher the number of words per utterance, the longer it takes to compose a single message. More interestingly, the messier the turn-taking the longer the mean editing time: more effort is required to repair miscomprehension due to confusion. Moreover, the higher the number of geo-reference of the kind 'label' or 'deictic' the shorter the time required, in average to compose a message. These techniques of naming icons and places on the shared workspace constituted the foundation of a proper linguistic strategy. Conversely, when this strategy was not in place, or the tool did not help establishing it, then the participants had to rely on a 'relative' geo-reference (e.g., "*On the right hand side of the building CE, ...just below*"), which took longer to compose.

Figure 7-12 represents the relation among the experimental variables visually as derived from the analysis reported in this chapter. The core finding to note here is that participants adapted to different communication constraints. In those situations in which they could use a tool supporting Explicit Referencing, they took advantage of the opportunity offered by the tool producing shorter

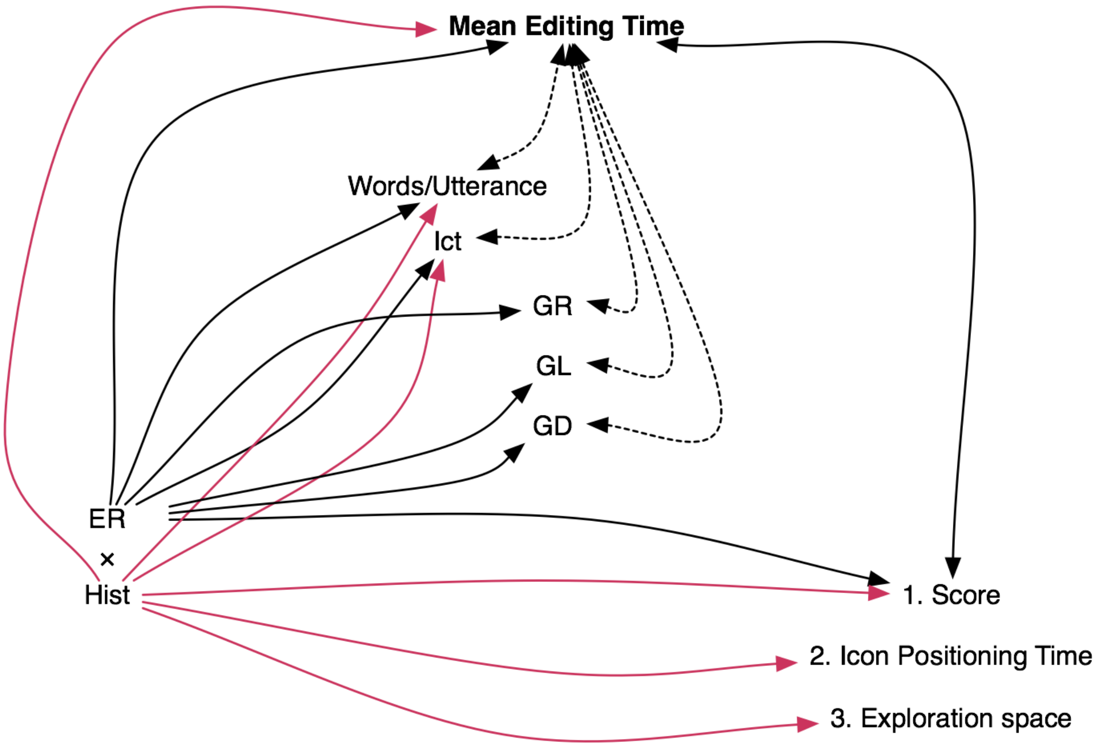


Figure 7-12: Significative interactions between the variables analyzed in this experiment. Lines represent significative regression or correlations between the connected variables. The message mean editing time is a mediation variable as it connects dependent and independent variables and its inclusion in the linear model removes the effect of the experimental manipulation on the obtained results

sentences. They employed the time they saved to compose messages about positioning the icons on the screen to figure out the logic of the game, or to test more solutions, subsequently achieving higher scores.

7.4.2 Implications

The results presented in this chapter have important implications for the design of systems that aim at supporting collaborative annotations of maps at distance. For such tasks providing links between the conversation and the shared landmarks over the map can be highly beneficial but only if the conversation does not fragment in providing this connection.

Additionally, I have developed this research looking at how Explicit Referencing might affect the positioning of objects on a shared map. Although maps are wonderfully rich artefacts that have specific cognitive implications on human cognition (Mark et al., 1999; Tversky & Lee, 1998), the findings of this work might be transferred to other domains of applications. For example, Google recently released a service called GoogleDocs[5] that allows editing a text collaboratively or

[5] See http://docs.google.com, last retrieved March 2008.

a working on a spreadsheet while chatting inline with the other participants. Depending on the complexity of the document or the number of participants, even pointing to specific parts of a text might become as difficult as routing objects on a map, and with even less spatial cues that a map can offer. I will expand these points in the concluding chapter.

Therefore, designers should carefully consider how to allocate conversation and task-context-support and what interaction mechanism is allowed between these two. Keeping them completely separated might not be optimal not just as combining them in the same visual space might have detrimental consequences as well.

7.5 Conclusions

This study presented a controlled experiment to compare the influence of different media in supporting collaborative work at a distance. I manipulated the experimental conditions in such a way so as to compare situations offering different support for what I defined as Explicit Referencing. My starting hypothesis was that tools offering this feature could better support collaborative tasks at a distance that require precise positioning of objects on a common display. Results indicate that team performance is improved by task referencing mechanisms unless the implementation of this mechanism is detrimental to the linearity of the conversation context (e.g., the message history), in which case, performance is reduced.

This study also demonstrated the importance of a linear message history for collaborative work at a distance. The manipulation of this factor affected many of the variables analyzed in this experiment. In particular, it had an impact on the three measures of performance analyzed above: the score, the time required to reach the first placement of the icon, and the exploration space. This factor was also deeply interrelated with the process variables that I studied: the average words per utterance produced by the participants, the index of complexity of the turn-taking and finally the mean editing time of the messages. In sum, the role of a linear message history in the collaboration mechanisms was equally important than that of Explicit Referencing.

From the distributed cognition perspective (E. Hutchins, 1995), I can summarize the obtained results by stating that each communication tool leads to a different organization of a distributed cognitive system (the user plus the tools). These different organizations produce different conversation styles employed by the collaborators. Tools like ShoutSpace or ExtremeChat, which were found less effective, led the participants to have a more formal interaction to compensate for the shortcomings of the tool they were using. Humans can cope with the limitations of a communication device up to a certain level, beyond which collaboration performance is affected.

Participants of the experiment took full advantage of the Explicit Referencing mechanism to reduce the number of words and the complexity of the spatial references needed to position the icons on the map. However, this analysis did not provide any answer about how participants actually looked at the appearance of these deictic gestures on their display. In particular, one can ask whether the the graphical signs of the anchorage points of the messages were perceived by the remote participants. Also, it would be interesting to understand whether the way they looked at the workspace had an effect on collaboration: was the way they looked at the display more similar when participants used ConcertChat compared to participants using MSN? Did they change strategy to place the icons because of the visual anchorage points produces by ER? These are the questions I seek to tackle in the next chapter.

Chapter 8

Indicating and Looking in Collaborative Work at Distance

This chapter extends the analysis of the experiment presented in chapter 7. It studies how the availability of Explicit Referencing affects the coordination of the eye movements of the participants.

8.1 Introduction

In chapter 3, I discussed how a remote deictic gesture might differ greatly from a gesture produced while face-to-face. In this chapter, I focus on two main reasons for this. First, deixis is always intertwined with eye gazing. A deictic gesture may be useless if not seen and acknowledged. While face-to-face, the conversant using a deictic gesture monitors with her gaze if the recipient has seen this communicative act. This is not possible at distance, at least not with commonly available technology. Second, some implementations of Explicit Referencing can break the linearity of the conversation, shifting from a time-organized flow (e.g., when in person, utterances follow temporally) into a space-organized flow (e.g., text messages might be accessed in a random order).

The previous chapter showed that supporting deixis in collaborative work at distance, and in the case of chat communication, might result in higher performance at a collaborative task. It also showed that the implementation of this mechanism might disturb the linearity of the conversation (e.g., messages distributed on the shared workspace as opposed to messages in the communication window linked to the shared map). This chapter presents a complementary analysis to these initial results, looking at whether Explicit References could influence the coordination of the collaborators' eye movements (e.g., participants looking at the same thing at the same time).

In this chapter, I use gaze recurrence, gaze coupling, and *cross-recurrence* as indications of this coordination. I therefore posed the following research question: *Q6, do collaborators using applications implementing Explicit Referencing look at the shared workspace in a more similar manner than collaborators using applications not supporting ER?* In other words, I was wondering whether the availability of virtually transferred deixis would have been acknowledged through gaze even without the presence of a visual feedback channel (see section 8.5). Additionally, I was interested in understanding whether more recurrent eye-gazes would lead to better performance.

Chapter 2 described how indicating is a communicative act used to attract the partner's attention and has therefore important implications in the way people reach a common understanding. It also explained how indicating comes intertwined with gaze in face-to-face conversations. Additionally chapter 3 discussed how deictic gestures are purely indicative acts, while gazing has the double nature of being both an indication device and a mechanism used to maintain a constant awareness of a conversation. The possibility of using these communicative devices has an impact on collaboration both when the participants are co-located and when they interact at a distance. Still, little is known about the interplay between these two mechanisms when people collaborate remotely. This chapter investigates whether supporting explicit referencing when participants solve a problem at distance influences the correspondence of the collaborators' eye movement. This analysis is also concerned about whether this degree of coupling of eye movements is helpful to the success of the collaboration. From the discussions reported in the previous chapters, I derived two hypotheses:

H1 The availability of explicit referencing mechanisms leads to a higher degree of gaze coupling (this concept will be defined in section 8.3 below);

H2 A higher degree of gaze coupling leads to higher performance.

I report the results of an analysis conducted using the data of the experiment described in the previous chapter. I show that, while a higher degree of eye coupling has a primal relationship with the collaboration performance, this gaze recurrence is not influenced by the availability of an Explicit Referencing mechanism (see section 8.4).

8.2 Method

This study extends the analysis of the experiment presented in the previous chapter. The data used to perform the cross-recurrence analysis explained at section 8.3 was collected using eye-tracking displays during the festival experiment reported in chapter 7. Therefore, the method employed was already presented in section 7.2.

I used the maximum score achieved during the 45 minutes of the task as the measure of performance. The process variables that I studied were:

1. The coupling of partners' eye movements, as explained in section 8.3 below;

2. The relation of fixation distributions, or the divergence of the two gaze distributions, as explained in section 8.3.3;

3. The use of linguistic deictic in the form of labels and prepositional phrases (e.g. "*Take that stage*", or "*Leave the icon where it is*"), as shown in section 8.4.4.

8.3 Cross-recurrence analysis

The goal of the study reported in this chapter is to understand whether participant `A` and `B` look at the same object at the same time. This is not easy to compute as what "same object" and "same time" means require a dynamic definition. Understanding which object the user is currently looking at is more complicated than just looking at the `x,y` coordinates of the eyes over the workspace. It needs to take into account the geometry of all the possible objects at sight. Similarly, the study of eye-movements requires a tolerance for delays (e.g. `B` might be looking the same object two seconds after `A`). While the former issue was here tackled with a simple radius of tolerance (often called geo-fence), the latter issue was tackled with the analysis described after.

To understand the relation between the eye movements of the speaker and the listener, I used *cross-recurrence analysis* (Eckmann et al., 1987). Cross-recurrence plots permit visualization and quantification of recurrent state patterns between two time series representing the evolution of dynamical systems. This is the technique that Richardson and Dale (2005) adopted in a listener's comprehension task. This analysis is useful as it can reveal the temporal dynamics of a data set without the limitation of making assumptions about its statistical nature. Figure 8-1, used by Richardson & Dale, 2005 , gives a graphical representation of this technique (p. 1050):

> Each diagonal on a cross-recurrence plot corresponds to a particular alignment of the speaker's and listener's eye-movement data with a particular lag time between them. A point is plotted along that diagonal whenever the speaker and listener's eye movements are recurrent–whenever their eyes are fixating the same object. Note that if the speaker and listener are not looking at any object at the same time (they were looking at blank spaces or off the screen or were blinking) this is not counted as recurrence.

On the left side of figure 8-1, the scarf plots of the speaker and the listener are aligned with no time lag. The periods counted as recurrence are shown in black in between these two linear plots,

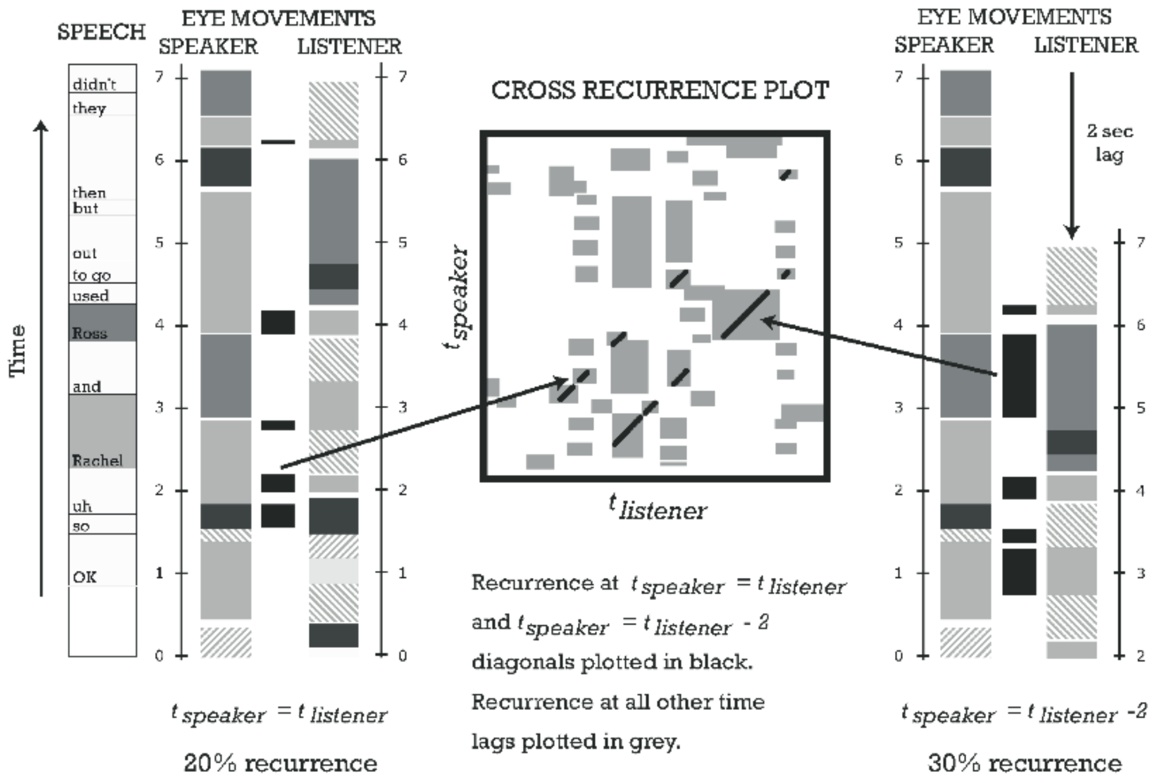


Figure 8-1: Scarf-plot and explanation of cross-recurrence analysis (from Richardson & Dale, 2005)

accounting for 20% of the time series. Conversely, on the right side of figure 8-1, the listener's eye movements are lagging behind the speaker by 2 seconds. Thus, there is a 30% recurrence between them. The recurrence analysis consists in calculating the recurrence between all such possible alignments.

8.3.1 Adaptations

While this method appeared to be a valid technique for analysing the eye-tracking data that I collected in the 'festival' experiment, my particular situation required a number of adaptations. First of all, the method of Richardson and Dale was taking into account discreet zones of interest in the shared visual space[1]. Conversely, the map was considered as a continuous space and therefore the recurrence in this situation did not mean being in the same discreet zone at the same time, but rather that the eye-movements of the first participant were within a certain pixel distance from those of the second participant and for a particular time interval.

Additionally, while Richardson and Dale analysed asymmetric interactions (e.g., one participant was speaking and the other listening), I had symmetrical interactions between the two

[1] The authors subdivided the screen space is six squares containing different visual stimuli. See `http://psych.ucsc.edu/eyethink/eye-chat.html` for a description of the experiment. Last retrieved March 2008.

participants (the two could be both emitter and receiver of a message). This resulted in different calculations to be done on the cross-recurrence matrix, as explained below.

Also, while they analyzed sequences of interactions of 5-10 minutes using a head-mounted eye-tracker, I analyzed interactions lasting over 45 minutes with an eye-tracking display. Although my system was less invasive, it had the side effect of loosing the tracking of the eyes if the user assumed an undesirable position in front of the display. During the task time, many participants become tired of sitting still and bent down over the table, thus provoking 'holes' of eye-tracking data in the collected dataset. Additionally, while they were typing their messages they sometimes looked at the keyboard in which case the tracker loose momentarily the position of the eyes. Therefore, the calculation of the cross-recurrence analysis had to take into account this missing data, a problem that Richardson and Dale did not have to solve.

Finally, I had to normalize the data as the registered cross-recurrence was dependent on the particular strategy that the participants chose. Working on a small portion of the screen increased the chance of two fixations to be considered recurrent, while working on a wider area decreased this level of recurrence by reducing the possibility of accidental overlaps. Therefore, I had to take this factor into consideration to compare the results of different experiments (as explained in section 8.4.1).

8.3.2 Procedure

Three steps were required to perform this analysis. *First step.* It was necessary to sample the gaze fixation data in order to obtain a continuous time series containing eye-gaze position every 200ms (A). Sampling points that fall during a fixation were simply assigned the position of the corresponding fixation. If no fixation was found for a given time, an interpolation was performed between the preceding and the next fixation, but only if they were separated by time less than 1000ms (B). Parameters (A) and (B) were chosen in order to have a sufficiently great (and comparable) number of fixation points for each experiment. A 'hole' in the dataset with a length shorter than one second could be due to movements of the eyes outside the screen area. So, if the fixations were temporally too distant, no data was taken between them. Moreover, I decided to reject all fixations falling outside of the map because I was looking at recurrence caused by the Explicit Referencing mechanism, which was acting only on the part of the screen displaying the map. Thus, depending on the eye-gaze tracker data quality and the ratio of map fixations, the resulting sampling contained between 10% and 50% of good points.

Second step. The next step was to compute the cross-recurrence matrix based on this sampling. This matrix is computed with the equation 8.1, which has been adapted from Eckmann et al., 1987 in order to ignore the missing sampling points:

$$CR_{i,j}(\varepsilon) = \begin{cases} \Theta(\varepsilon - \|\vec{x_i} - \vec{y_j}\|) & \text{if eye-tracking data is available,} \\ -1 & \text{if eye-tracking data not available.} \end{cases} \tag{8.1}$$

Where i and j are the number of the sampling points, and $\vec{x}$ and $\vec{y}$ are the sampled fixation data for the first and the second participant, respectively. Also, Θ is a step function which returns 1 when its argument is positive and 0 when it is negative (see formula 8.2 below).

$$\Theta(z) = \begin{cases} 1 & \text{if z > 0,} \\ 0 & \text{if z < 0.} \end{cases} \tag{8.2}$$

In formula 8.1, epsilon represents the threshold under which two fixation points are considered to be recurrent. There is no generally valid method to set this threshold, which is very dependent on the system under consideration (e.g., the size of the objects in the shared workspace). I chose to take 30 pixels, a measure that is slightly larger than the eye-gaze tracker accuracy when the user sits at 60 cm from the screen and smaller than most of the polygons composing the map used in the experiment.

Third step. these cross-recurrence matrices were used to compute the recurrence rate at different time lags. Indeed, if I compute the ratio of recurrence points along the diagonals in these matrices, they correspond to the recurrence rate at a given time lag (see figure 8-2), the identity diagonal being the recurrence with no time lag. When missing data was present (e.g., $CR_{i,j} = -1$), it was simply ignored in the computation of the recurrence ratio, which had the effect of increasing noise for the experiments with too few good sampling points. From these values, I plotted the recurrence rate for every time lag between 0 minutes and +5 minutes (see figure 8-3).

8.3.3 Relation of fixation distributions

The resulting graphs showed some incoherence: the randomized average recurrence was different than 0; it was different across pairs; and even across those conducted under the same experimental condition. To understand the reasons for this variability, I analyzed the cumulative spatial distribution of the sampled eye-gaze points. In order to achieve this, I computed a cumulative distribution of fixations over the shared workspace (the points looked at during the whole task) for each participant by subdividing the map area in small cells and by counting the number of fixation points falling in each cell. Then, I computed a distance measure between these two distributions using a discrete version of the Kullback–Leibler divergence (KL in short). This is a non-commutative measure of the difference between two probability distributions `P` (in this context, the eye-movements of participant 1) and `Q` (in this context, the eye-movements of participant

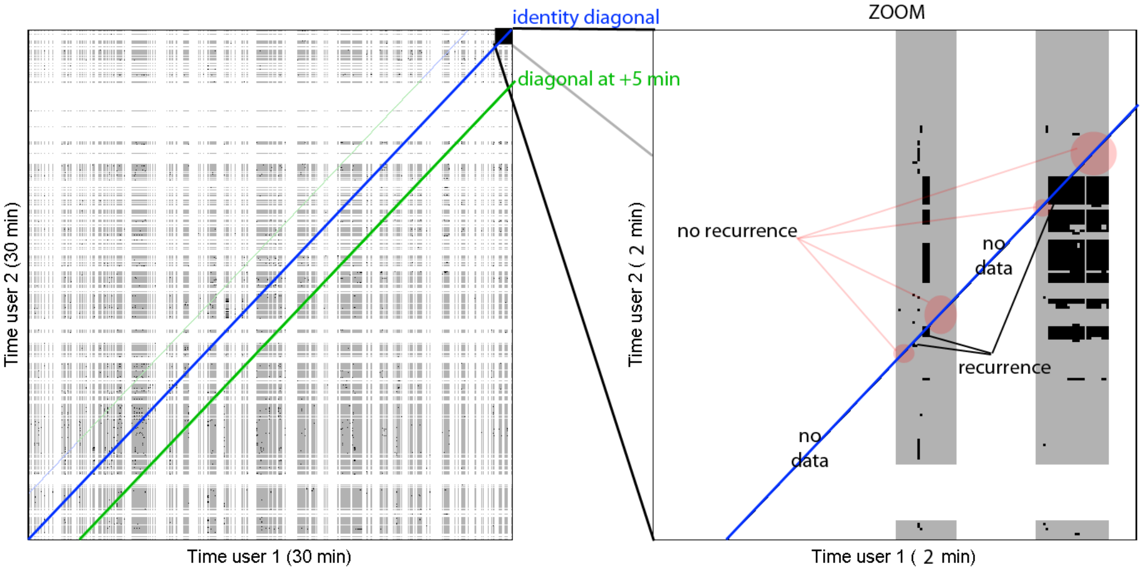


Figure 8-2: Example cross-recurrence plot of the eye movements of the participants: on the left hand side the matrix plotted for a 30 minutes period. The central diagonal corresponds to a lag of 0 seconds. A segment of 2 minutes is enlarged on the right hand side. Gray areas represent times for which I have readings from both eye-trackers

2)[2] (Kullback & Leibler, 1951). The results of this analysis are presented at section 8.4.1 below.

8.3.4 Randomized level of eye-movements

In order to analyze the curves generated by the cross-recurrence plot explained above, the curves had to be compared with a baseline distribution. This was created by shuffling the temporal order of fixations generated by a certain pair. This randomized series was calculated for each trial, and served as a baseline of looking “at chance” at any given point in time, but with the same overall distribution of looks to the map as in the real collaborations. This measure was used in the analysis reported in section 8.4.2.

8.4 Results

Of the original 60 experiments, I discarded 9 recordings of pairs that, for technical problems, were missing logs. For each of the remaining experiments I computed the number of fixations being sampled. This measure was used to further exclude 18 experiments, which had less than a thousand fixations falling on the map during the 45 minutes of the task time. I finally generated cross-recurrence plots for the remaining 33 experiments (MSN chat: 5; ConcertChat: 7; ShoutSpace: 9; ExtremeChat: 12).

[2] See `http://en.wikipedia.org/wiki/Kullback-Leibler_divergence`, last retrieved March 2008.

8.4.1 Relation between task strategy and gaze recurrence

To measure this relation, I computed a linear regression between the Kullback–Leibler divergence of the fixations-points distributions of the pair and the maximum recurrence. The regression of the maximum recurrence was a good fit, describing 43.8% of the max-recurrence variance (R^2_{adj}= 42.0%). The overall relationship was statistically significant (F[1,32]=24.93, p<.001). The Kullback–Leibler score was negatively related with the maximum cross-recurrence (β_{std}=-.66, p<.001).

The analysis revealed a significant relation between the KL divergence of the eye-gaze points distribution of the two participants and the recurrence rate of the randomized for the same experiment. So, I concluded that the difference between randomized recurrence rate of the experiments were due to different strategies employed by the participants in exploring the map. Participants pairs working in a smaller portion of the map could have an higher chance to be looking at the same points of the map compared to participant pairs working on a larger portion of the campus plan. Thus, in order to be able to compare the recurrence rates between different experiments, it was necessary to suppress this intra-experiment effect. This was accomplished by simply subtracting from each experiment's recurrence distribution the average of the randomized gaze recurrence for the same experiment.

8.4.2 Explicit Referencing and the relation between collaborators' eye movements

H1, The availability of explicit referencing mechanisms leads to a higher degree of gaze coupling.

The initial question that I addressed was what experimental condition produced the most recurrence between the collaborators' eye movement. Figure 8-3 shows the average cross recurrence, corrected with the randomized level (see 8.3.4 and 8.4.1), at different time lags and for each experimental condition.

The differences between the experimental conditions were supported by a 2 (Hist–noHist) × 2 (ER–noER) × 91 (lag times) mixed-effects analysis of variance (ANOVA) (lag as a repeated measure factor) that denied a significant effect of the availability of a liner history, F[1, 29]=.038, p>0.1, ns. The same analysis also denied a significant effect of the availability of Explicit Referencing, F[1, 29]=.00, p>0.1, ns. This result was not consistent with H1, which was predicting a higher recurrence rate for experimental conditions supporting Explicit Referencing.

Figure 8-3 also shows a baseline distribution where I calculated the recurrence of eye movements of participants where I shuffled the temporal order of the eye-movement sequence, offering a comparison of random looks (gray curve oscillating around '0%' recurrence ratio)[3]. This contrast shows that the eye movements of the two collaborators are linked within a particular temporal

[3] Of course, as explained at section 8.3.4 and 8.4.1, these randomized curves were originally placed at different levels and then equalized around 0% recurrence level.

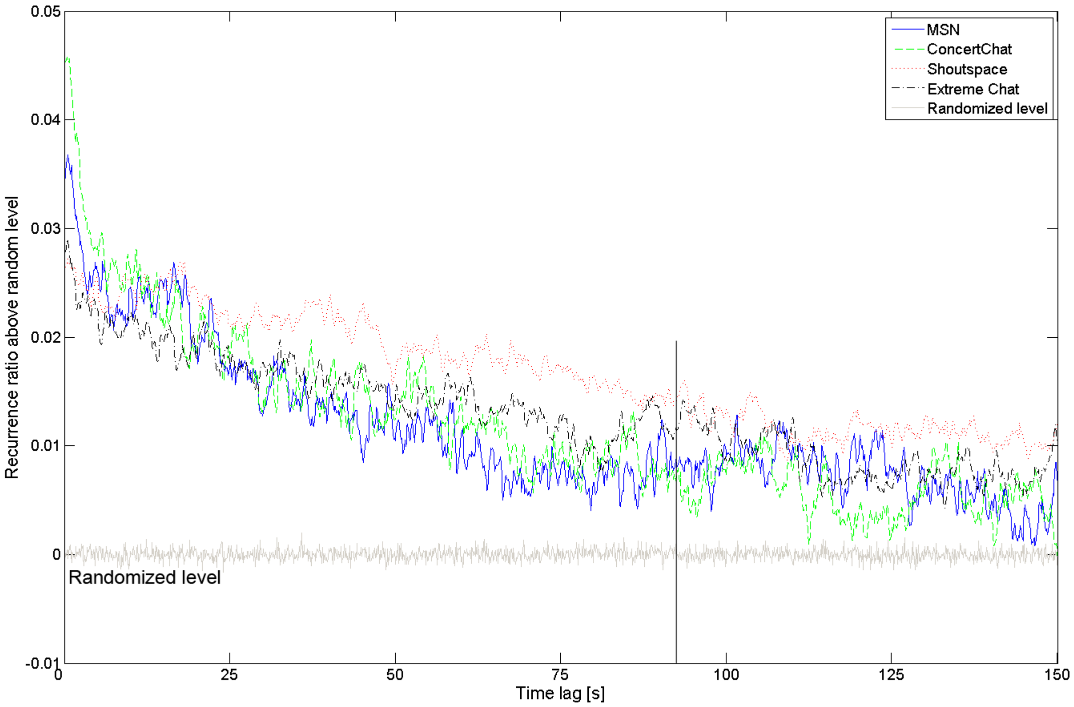


Figure 8-3: Average cross-recurrence for each experimental conditions. The curves have been smoothed for readability. The line at about 90 seconds marks the peak area of the curve

window: between 0 seconds and 1 minute and 30 seconds, the participants are likely to be looking at the same thing at above chance level. The maximum recurrence for all the curves is around 0 seconds.

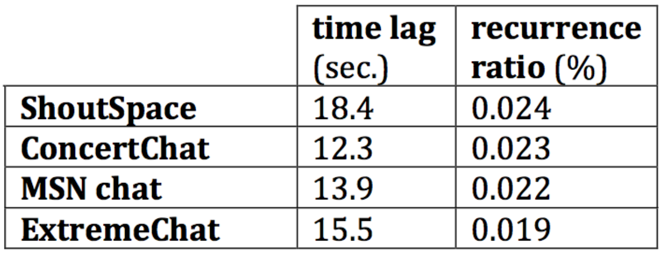
The differences between the real pair and a randomized pair were supported by a 2 (real–randomized) × 91 (lag times) mixed-effects analysis of variance (ANOVA) (lag as a repeated measure factor) that revealed a significant effect of pair type, F[1,66]=39.46, p<.001, and a main effect of lag, F[90,5940]=5.54, p<.001. There was also a significant interaction between the factors, F[90,5940]=5.93, p<.001. This implies that real pairs were looking at the same things at the same time, or with a constant lag, at above chance level.

I performed the presented analysis increasing the time lag between the participants up to 10 minutes. However, the most interesting part of the curve was between 0 and 150 seconds. Figure 8-4 present visually the peaks of the curves. For enhancing the readability of the cross-recurrence plot, the graph was smoothed by applying a low-pass filter function[4]. The maximum values of these smoothed curves are summarized in table 8.1.

To summarize: these results denied an effect of the availability of the ER mechanism on the amount of gaze coupling reached by the pair. The comparison with the baseline distribution of

[4] A low-pass filter is a function that attenuates (reduces the amplitude of) signals with frequencies higher than the cutoff frequency. See `http://en.wikipedia.org/wiki/Low-pass_filter`, last retrieved March 2008.

Table 8.1: Summary of the maximum values reached by the smoothed cross-recurrence curves presented in figure 8-4



	time lag (sec.)	**recurrence ratio (%)**
ShoutSpace	18.4	0.024
ConcertChat	12.3	0.023
MSN chat	13.9	0.022
ExtremeChat	15.5	0.019

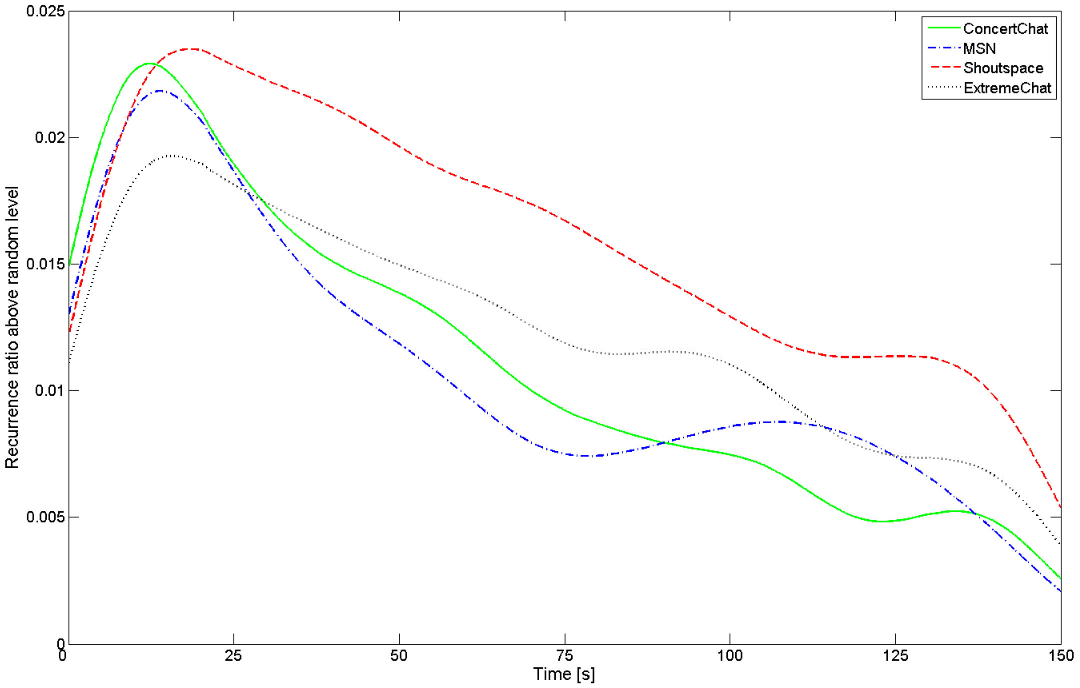


Figure 8-4: Average cross-recurrence for each experimental conditions between 0 and 150 seconds. The curves have been smoothed for readability with a low-pass filter

random looks demonstrates that participants' gaze movements are coupled. The maximum of gaze recurrence was reached in average with a lag of 15 seconds. In other words, after 15 seconds one of the participants was looking at some points of the map, the other participant was also looking at the same points.

8.4.3 Gaze recurrence and pair's performance

H2, A higher degree of gaze coupling leads to higher performance.

Was the degree of coupling of the participants' eye movements related to the maximum score obtained by the pair? To answer this question I measured two characteristics of the peak of each experiment's recurrence curve: the maximum recurrence and the average recurrence between 0 seconds and +1 minute (many messages took one minute to be composed). I computed a linear regression of the score in relation to these two measures. The regression of the maximum recurrence was a poor fit, describing only 27.4% of the score variance (R^2_{adj}= 25.2%), but the overall relationship was statistically significant (F[1,32]=12.09, p<.001). The pair score was positively related with the maximum cross-recurrence, increasing by 1.56 points for every extra percent of recurrence (β_{std}=.52, p<.001). This findings was consistent with H2, which predicted a higher

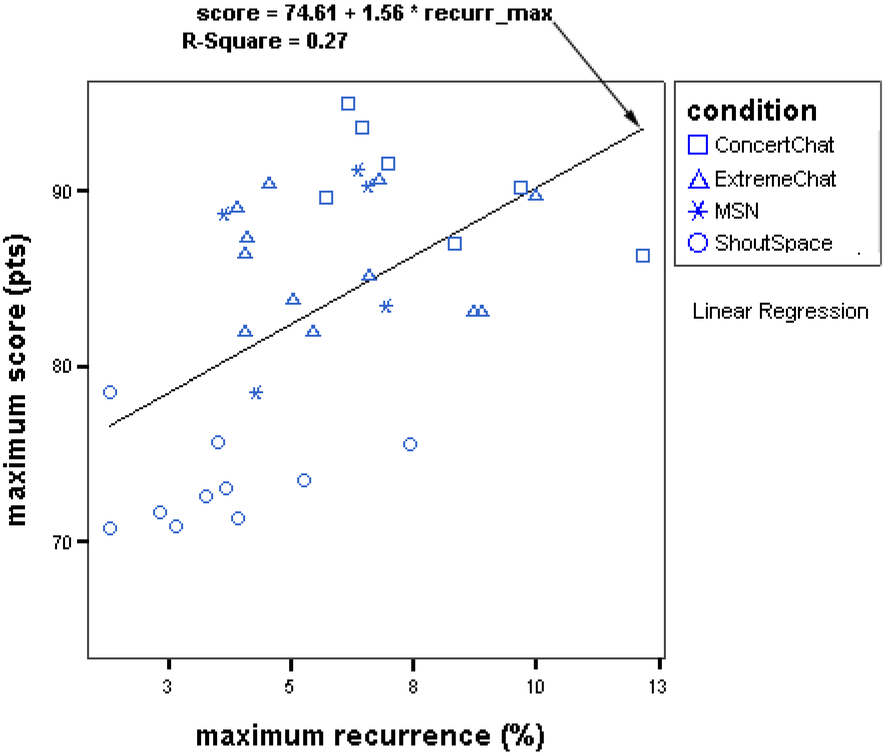


Figure 8-5: Scatter plot of the maximum recurrence and the maximum score with the regression line

score for pairs with an higher gaze recurrence. This implies that the more the gaze movements of the collaborators are coupled the higher performance may reach their interaction. Figure 8-5 shows the relation between the maximum recurrence and the score. The average recurrence calculated between 0 seconds and +1 minute was also positively related with the score (R^2_{adj}=.18, β_{std}=.45, p<.05).

As the presence of the Explicit Referencing mechanism was not found responsible for an increase of gaze coupling other process variables were analyzed to account for empirical differences among the experimental conditions.

8.4.4 Use of linguistic deixis and gaze recurrence

As the task required precise positioning on the shared map, I counted the number of deictic expressions used in the conversation (e.g., *"I placed the second concert here"*, or *''move your icon there"*). This measure was related to the use of explicit or implicit referencing in the messages. On average, deictic expressions were employed more frequently in ConcertChat or ShoutSpace, as reported in the previous chapter.

As I could not find a direct relation between the experimental conditions and the recurrence of the eye movements, I looked at the relation of this indirect measure of deixis and the maximum gaze recurrence. The maximum recurrence was positively related with the number of deictic expressions used in the messages (R^2_{adj}=.11, β_{std}=.37, p<.05). I did not, however, find a significant relation of the number of deictic expressions with the score (R^2_{adj}=-.02, β_{std}=-.02, p>0.1, ns).

8.5 Discussion

Q6, do collaborators using applications implementing Explicit Referencing look at the shared workspace in a more similar manner than collaborators using applications not supporting ER?

Eye movements of collaborators are linked. When I compared the distributions of the different conditions with the distributions of randomly matched pairs, I observed that the recurrence peak was not present, thus suggesting that the visual attention of the pair was indeed coupled (see figure 8-3). The form of the peak gives many cues on the differences between the conditions on the way participants interacted. It is possible to see that the curves reach their right after 0 seconds (15 seconds in the curves smoothed with a low-pass filter, see figure 8-4). As participants often contributed symmetrically to the task, the recurrence distribution was coherently centered on 0 seconds. However, the recurrence curves presented many local maxima. Smaller peaks at further distance might be due to different categories of messages with longer editing times (e.g., utterances containing positioning indications in relation to other elements, which would have taken more time to encode).

H1, The availability of explicit referencing mechanisms leads to a higher degree of gaze coupling.
I performed the study reported in this chapter as I wanted to understand whether the availability of Explicit Referencing had an impact on the way people looked at the shared workspace. To this end, this chapter reports the following finding: **the manipulation of Explicit Referencing did not influence the cross-recurrence ratio**, the statistics computed to answer the research question, which measures the degree of visual overlap of the pair for a particular time lag. Therefore, H1 was not verified. While the manipulation of the availability of Explicit Referencing had an effect on the maximum score reached by the pair (see chapter 7), this was not the case for the maximum recurrence of the eye movements registered for each team. This means that pairs interacting at distance over a shared map and communicating with a standard chat application looked at the same areas of the map simultaneously, or with a constant lag, with the same frequency of pairs using a chat application implementing an Explicit Referencing mechanism.

H2, A higher degree of gaze coupling leads to higher performance.
This study produced a second important finding. The regression reported in section 8.4.3 shows that the degree of coupling of eye movements was related to the performance obtained by the pair. This implies that pairs that look more often at the same thing at the same time, or with a constant lag, obtain higher scores, even when people cannot observe each other's faces. This result extends the findings of Richardson and Dale (2005). While they found a correspondence between the degree of cross-recurrence and the scores of a post-hoc comprehension questionnaire, I measured the relation of cross-recurrence and performance in a collaborative problem-solving task. Therefore, H2 was verified.

Gazing was a personal and self-directed activity because each collaborator knew that her partner could not see where she was looking. When a participant wanted to invite the partner to look at the same point she was looking at, she used the Explicit Referencing mechanism to circumscribe the referential domain of a message. However, the emitter of one of such enriched messages had no direct indication that these 'acts' were subsequently observed by the conversant.

While I did find a relationship between the degree of coupling of the eye movements and the task score, I did not find the same connection between the use of linguistic deixis and task performance, as described in 8.4.4 (figure 8-6 summarizes the effects found). In the experiment, when participants used linguistic deictic expressions, they affected their partner's attention, and therefore the places she looked at on the shared display, more than when participants used Explicit Referencing to convey deixis. This suggests that the degree of eye coupling has a primary relationship with task performance, while the frequency of use of deictic gestures, expressed through visual links to the shared display or with text only, has an indirect relation with the collaboration outcomes.

To put this second finding simply, the more gaze movements of the participants were coupled,

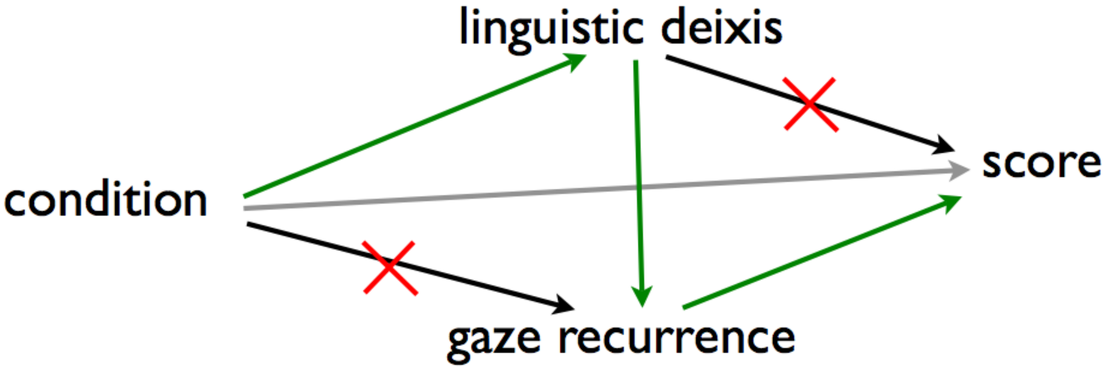


Figure 8-6: Significative interactions between the variables analyzed in this experiment. Lines represent significative regression or correlations between the connected variables

the better they performed the task. The reader might be lead to think that this is a perfectly obvious result, which is largely supported by common sense, and that as a concept this underpins most of the work claiming the need for a shared visual space in supporting collaborative work at a distance. However, taken together, the results that I presented support the idea, which is not common in the literature, that this coordination of sight was better achieved here through the linguistic channel (the chat application) than through the visual channel (the map). The reasons why this was the case might be numerous. Here, I suggest three. First, as stated before, the implementations of Explicit Referencing that I used in the experiment did not allow any explicit acknowledgment by the recipient of a message. Second, the production of a message containing a reference to the plan might not have been as effective in capturing the attention of the partner as a linguistic message directed to the management of the interaction. Third, this result might be a reflection of the fact that the chat tools tested were particularly poor at constructing Explicit Referencing, although I did register an effect of ER on performance, as reported in chapter 7. The results provided by the analysis presented in this work do not suggest a conclusive explanation on this matter.

As these results stand, they show that pairs that communicated using linguistic 'shortcuts' as deictic expressions, were able to capture their partner's gaze more efficiently, as demonstrated by the higher degree of gaze coupling. This happened independently of the tool they were using to communicate and resulted in higher scores reached by the pair. We can therefore speculate that there exist a relation between the eye-movements of a person and her attentional focus. However, more research is needed to understand the nature of this relation.

8.6 Qualitative analysis of the eye-movements

The cross-recurrence analysis presented in the previous section offered a macroscopic picture of how the participants looked at the shared workspace over time. This technique showed how, on average, collaborators tended to look at the same icons or the same zones of the map. However, this analysis could not give a clear picture of what happened at microscopic level. The interest in performing an analysis at the level of a single message lies in understanding whether the production of a message containing a particular kind of geo-reference would have influenced the recipient and the emitter to look more closely to the same parts of the screen. In order to test this, a visualization of the eye-movements produced during a specific linguistic sequence was built. This visualization merges the users' communication with the users' interaction with the shared artifacts (icons and map). These interaction maps were inspired by similar work in the field of visualization of comunicative exchanges (Donath et al., 1999; Viégas et al., 2004).

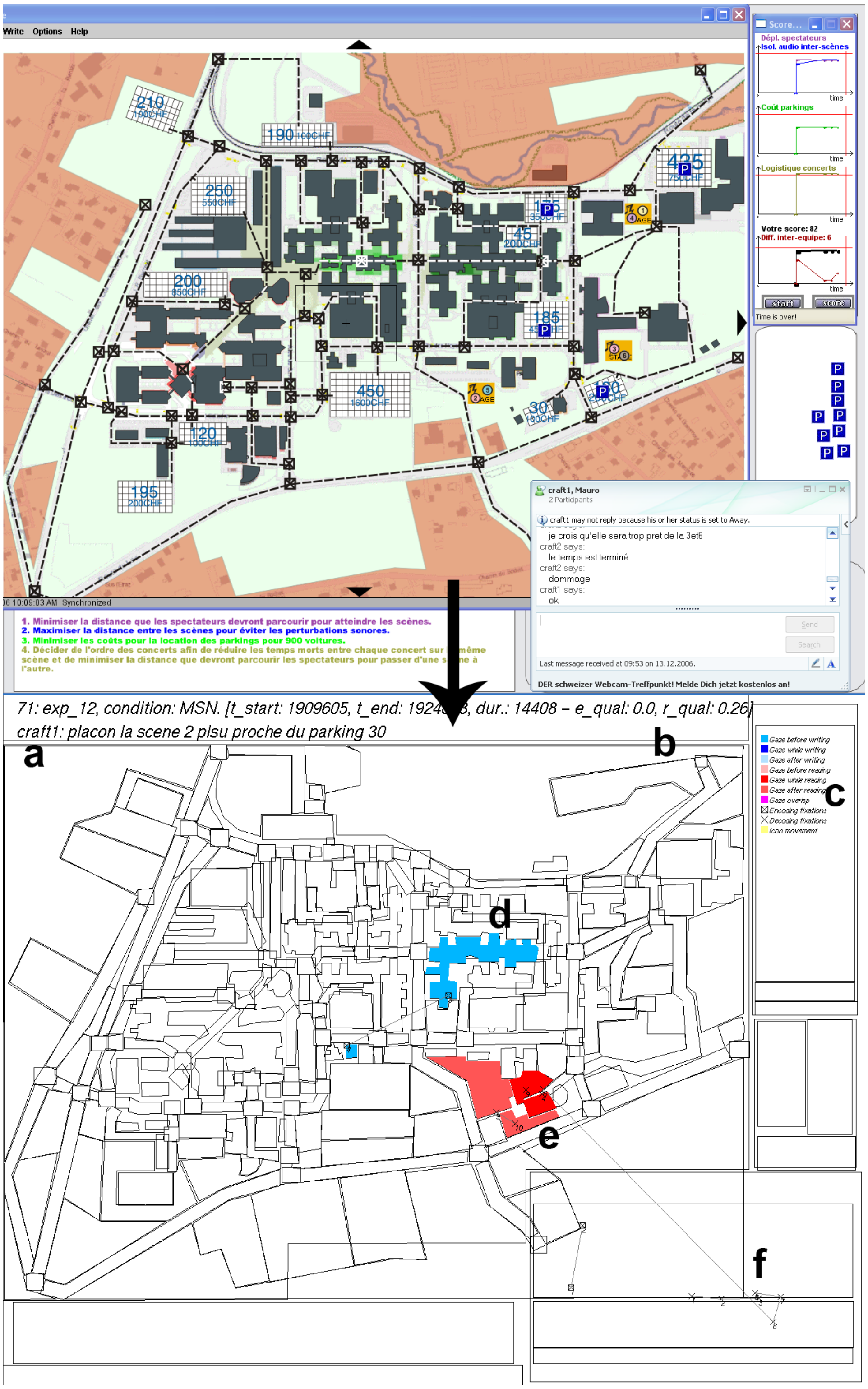
One of the key elements of building a visualization of a multi-dimensional data array is the challenge of providing a sense of time progression of the events. Tufte offers great guidelines on how to build narrative graphics of space and time (Tufte, 2001, p. 40):

> An especially effective device for enhancing the explanatory power of time-series displays is to add spatial dimensions to the design of the graphic, so that the data are moving over space (in two or three dimensions) as well as over time.

In the context of this work, I had to visualize the time series of the eye-movements of two participants, in relation to the shared workspace that was used during the task and to the linguistic events that were produced. The composition of the message was used as the keyframe according to which the other data should have been treated. Therefore, the edition time of a certain message was used to filter the gaze events that were relevant for the emitter of the message. Additionally, I used the interval between the time of reception of the message and the time of a subsequent action as the window to extract the events that were relevant for the recipient of the message.

A second choice taken was that of preserving the spatial relations of the workspace in order to give a sense of the spatial areas that attracted the users' attention during the interaction. The workspace was subdivided in polygons representing the sub-areas that composed each zone. For instance, the map of the campus was defined by many polygons: buildings, streets, nodes, plazas, parking lots, etc. The chat application was also defined by different sub-zones: the message composition pane, and the history of previous messages (where present). Finally, the users' fixations over the workspace were represented as numbered dots connected by traits to give a sense of sequence to the series (see (e, f) of figure 8-7).

From qualitative observations of the video recordings of the users' interaction, I learned that relevant fixations for a certain message were produced not only during the composition of the



1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les temps morts entre chaque concert sur la même scène et de minimiser la distance que devront parcourir les spectateurs pour passer d'une scène à l'autre.

craft1 may not reply because his or her status is set to Away.

je crois qu'elle sera trop pret de la 3et6

craft2 says:
le temps est terminé

craft2 says:
dommage

craft1 says:
ok

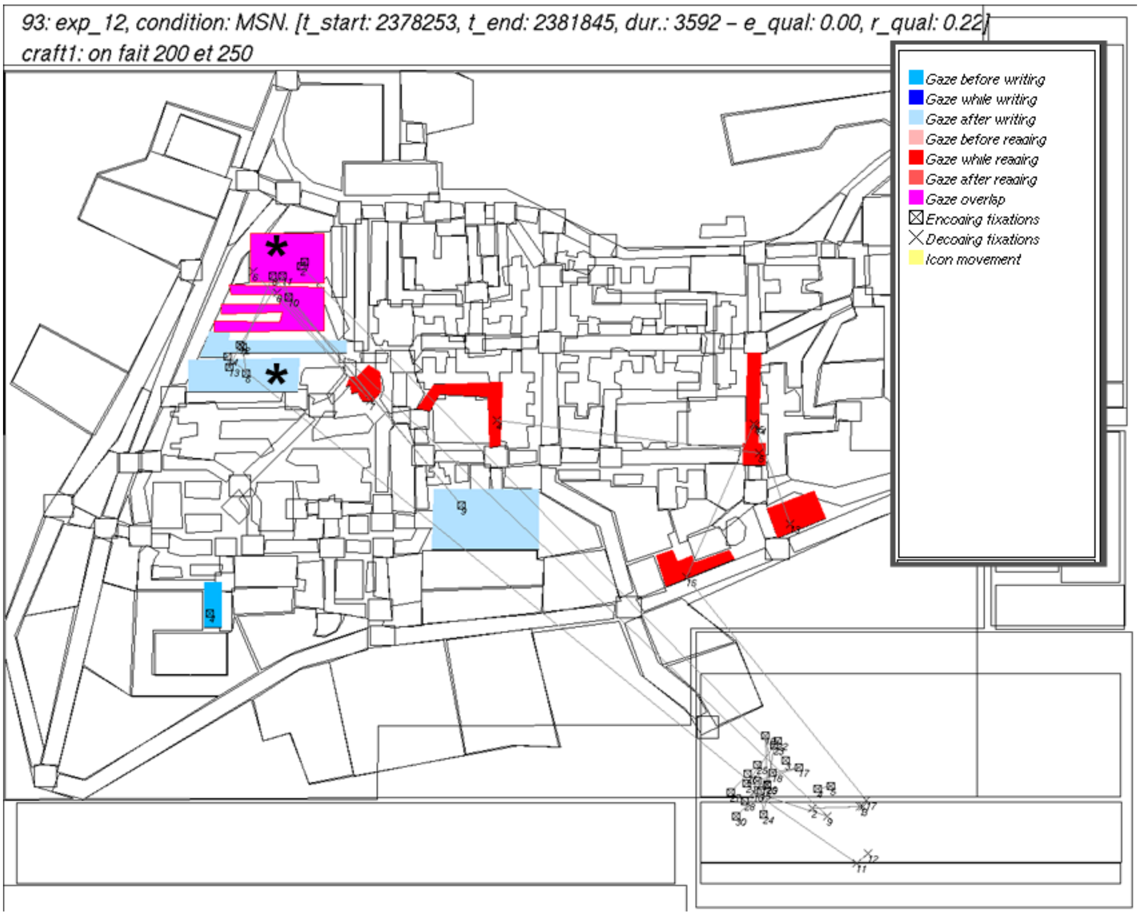
Last message received at 09:53 on 13.12.2006.

DER schweizer Webcam-Treffpunkt! Melde Dich jetzt kostenlos an!

71: exp_12, condition: MSN. [t_start: 1909605, t_end: 1924013, dur.: 14408 – e_qual: 0.0, r_qual: 0.26]

craft1: placon la scene 2 plsu proche du parking 30

Figure 8-7: Map visualization of the users' interaction (left) and comparison with the original screen capture (right). (a) message on which the data was filtered; (b) quality of the eye-tracking data for the composition (e-qual) and retrieval (r-qual); (c) color key; (d) encoding fixations; (e) decoding fixations; and (f) fixations in the message window

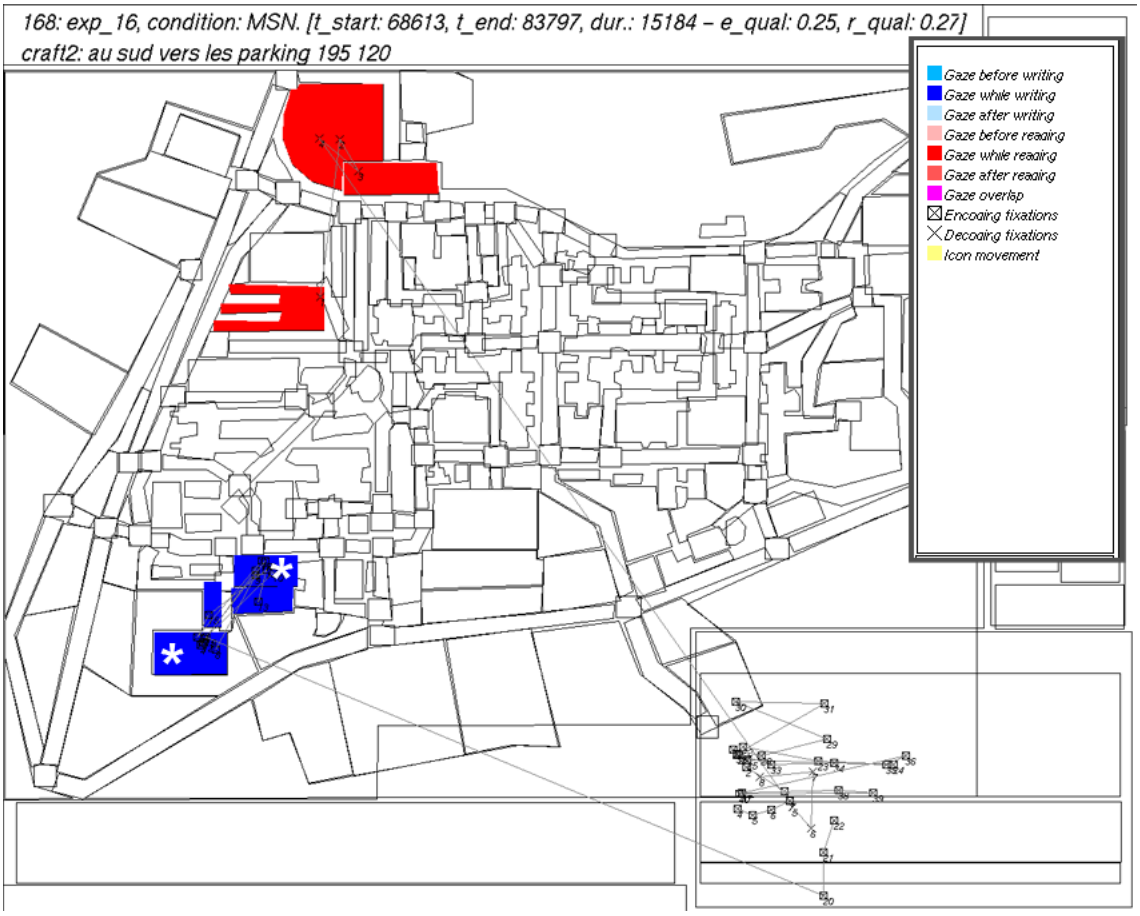


93: exp_12, condition: MSN. [t_start: 2378253, t_end: 2381845, dur.: 3592 – e_qual: 0.00, r_qual: 0.22]
craft1: on fait 200 et 250

Figure 8-8: Example of good gaze coordination. Asterisks indicate the polygons named in the message (target polygones). The color key has been enhanced for readability

message but also right before the start of the editing time and rarely after the message was sent to the collaborator. Similarly, I learned that the recipient of a message was producing fixations that could have been considered relevant for a certain message both during the time the message was read and after (e.g., when the recipient's eye last left the message history pane). Therefore, I assigned color codes to each polygon of the map over which either the emitter and the recipient of a message fixated, while and after the message was composed and read (see (c) of figure 8-7). Figure 8-7 summarizes my choices and presents the visualization that I built to perform qualitative analysis of microscopic episodes of coordination implying gaze and linguistic events.

The analysis of the visualizations helped me to ascertain that in a majority of cases, the emitter of a message fixated on the polygon she talked about in the message. When there was good coordination between the two participants this was usually easily visible through overlaps of fixations in the same polygon. An illustrative example of this kind of situations is given by figure 8-8, where the emitter typed the message: *"in fact 200 and 250"*. She was referring to the

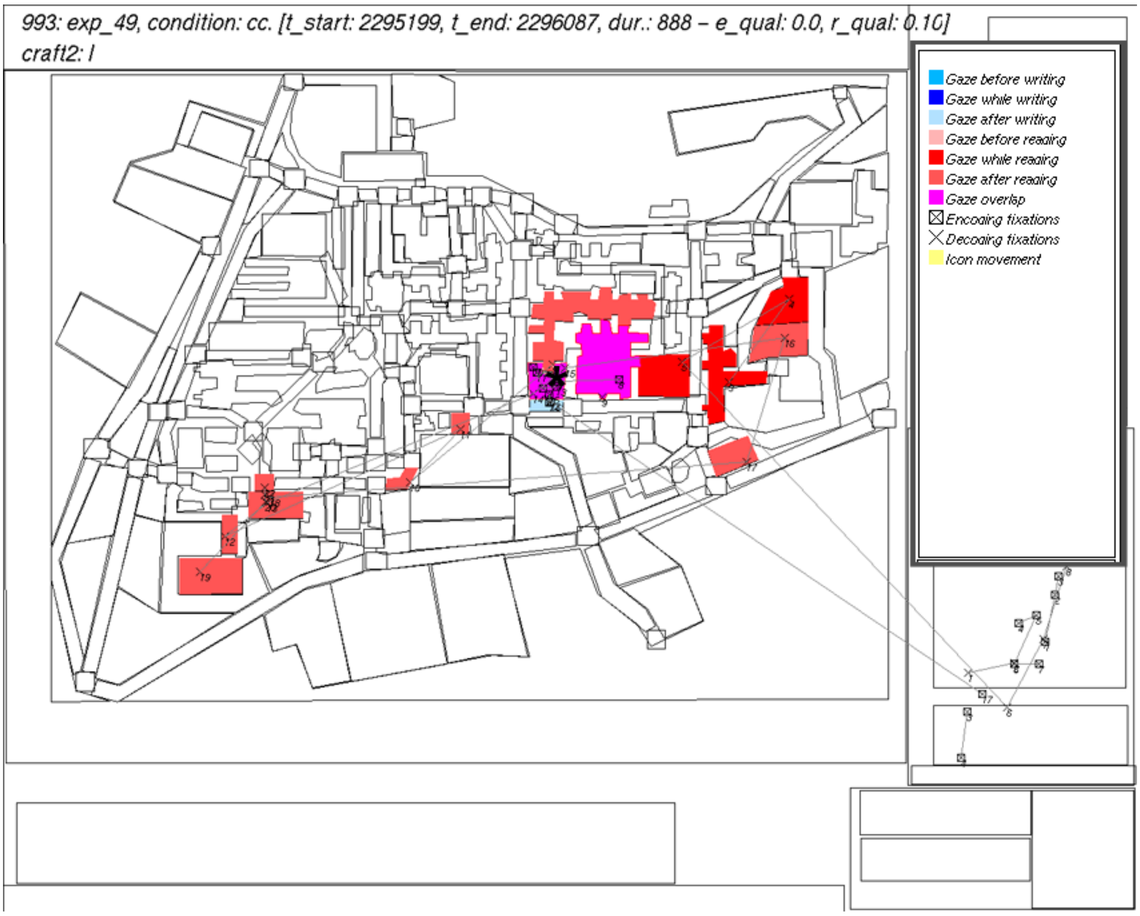


168: exp_16, condition: MSN. [t_start: 68613, t_end: 83797, dur.: 15184 – e_qual: 0.25, r_qual: 0.27]
craft2: au sud vers les parking 195 120

Figure 8-9: Example of 'bad' gaze coordination. Asterisks indicate the polygons named in the message (target polygones). The color key has been enhanced for readability

parking lots of 250 and 200 places which were available on the north-west side of the campus (the corresponding polygons are marked with asterisks in figure 8-8). While she was writing the message, the emitter made fixations on the map in the correspondence of the parking lots labeled 'P250' and 'P200' and on the buildings in between. Then she sent the message. While reading the message, the recipient fixates on 'P250' and on the building just below it. He also fixates on a number of other points that were not directly related to the message. Perhaps, he needed to check whether there was only one P250 on the map, or maybe he was looking for other possible spots that could be suitable for the placement of the icon that was under discussion.

In the next example, represented in figure 8-9, I present the opposite situation to the one presented above. Here, the emitter of the message types the following: *"on the south towards parking 195 120"*. While composing this message, she made a number of fixations on the two polygons named in the message and on those in between. While the recipient of the message was reading its content, he fixated on the topmost part of the map, which was not connected to the



993: exp_49, condition: cc. [t_start: 2295199, t_end: 2296087, dur.: 888 – e_qual: 0.0, r_qual: 0.10]
craft2: I

Figure 8-10: Example of gaze coordination in ConcertChat condition. The asterisk indicate the polygons named in the message (target polygon). The color key has been enhanced for readability

polygons named by her partner. Additionally, he did not make a single fixation in the area that the emitter was talking about in her message. So, given the information I had at my disposal, I could not establish whether this behavior was intentional (I understand what you are saying but I choose to look to something else) or due to miscomprehension between the two collaborators (I look at the wrong parking as I think you are mentioning this one).

Finally, in the last example reported in figure 8-10, I present a case showing how participants in the ConcertChat condition took full advantage of the Explicit Referencing mechanism. Here, the emitter of the message types only one number in her message: *"1"*. However the message contained a reference to a particular polygon of the shared workspace (marked with an asterisk in figure 8-10). Therefore, the recipient of the message could easily understand *'where' the other was talking about*. This is demonstrated by the gaze overlap in the region around the named polygon. Additionally, the recipient of the message fixated on a wide number of other polygons, mainly parking lots. As conjectured earlier, perhaps he was thinking about a different strategy that he wanted to suggest to her partner in the next utterance. A final note on ConcertChat concerns the fact that the recipient of the message might fixate on the target polygon even before reading the content of the message. In fact, as the message is delivered, he can immediately see the anchor point. Therefore, it may happen that he fixates on the reference before entering the message box to read the content of the message.

The analysis performed with the visualizations presented in this section helped me to understand that good coordination in the festival task required also good gaze coordination. When participants were in the need of coordinating the placement of icons on the map they coordinated by sharing some reference points. For instance, they could say: *"Below P400"*, meaning that one of the stages could be placed in the empty spot below the parking lot with a capacity of hosting four-hundred cars. In the majority of the situations that I could observe with the visualization I built, emitters in those situations would have fixed on the reference polygon to make sure they were encoding the spatial relation in the correct way. Also, recipients of such messages containing references would have fixed them to the reference polygon to ensure of the correct decoding of the meaning expressed by their partner. I could also see situations in which this matching of the gaze movements with the places expressed in the message was not effective (e.g., participants looking at different polygons). This was usually associated with episodes of chat confusion, like intertwined turn taking.

This analysis gave me the idea that I could actually use the distance between the fixations of the first participant and those of the second participant in relation to the reference polygon as a measure of understanding between the collaborators. A shorter distance between the fixations of the two participants to the reference polygon could have been a symptom of good understanding, while wider distances might have indicated a less good coordination between the partners. Of course, the test of this hypothesis required the definition of a continuous measure of distance between the fixations of the participants and the reference polygons than the simplistic coloring that I employed in the visualizations presented. The technique I presented in this section is depended on the size of the polygons composing the map: polygons with a bigger area would give higher chance of overlaps than those with a smaller area. Additionally, such method should take into account extra fixations on the plan that were not directly related with the reference points indicated in the current message. I address these issues and I verify the hypothesis that I described above in the next chapter.

8.7 Conclusions

This chapter presents a study of gaze coupling for remote collaborative tasks using different forms of chat systems. It compares the differences between support for Explicit Referencing and a linear chat history. The study did not find a link between the use of Explicit Referencing mechanisms and gaze coupling but found a correlation between gaze coupling and task performance.

The analysis presents a second important finding: a relation between the frequency of linguistic deixis and the gaze recurrence. While further evidence is necessary to provide a sound explanation of this result, I hypothesized that in the presented experiment, the coordination of gazes was better achieved through the linguistic channel than through the visual channel supported by the Explicit Referencing mechanism. If these findings could be confirmed by further research, it would open a discussion on whether designers should strive to give more support to the visual side of interfaces aimed at supporting collaborative work at a distance. In the described experiment, participants achieved better coordination of sight using utterances than by using the annotation tool provided by the interfaces supporting ER.

Designers of systems aimed at sustaining collaborative work at distance should carefully consider how to coordinate the focus of attention of collaboration partners. This might be achieved indirectly through the disambiguation of context offered by Explicit Referencing or more directly through the visualization of the mutual focus of attention, namely the concentration of eye fixations on the shared workspace (this is often referred to as *gaze awareness*). However, the results presented here caution against to dissociating these two communication mechanisms as they interact in complex ways.

Finally, the qualitative analysis of gaze movements presented in this chapter provided evidence that emitters of a message containing a reference on the map used gaze to check the correctness of the information they were sharing through the chat tool. Similarly, the recipient of the message containing the reference on the plan gazed over the points mentioned in the messages to understand what the emitter was talking about. In specific situations, this was not the case as the recipient looked at polygons which were distant from the polygons named by the emitter. This qualitative observation raised the hypothesis that, in these latter circumstances, the recipient could have misunderstood the message (or he was thinking to something else). Additionally, this study gave me the idea that the behavior of participants in checking the right polygons during interaction could have been used to build a computational support for automatically detecting miscomprehension between remote collaborators. The next chapter will focus on this issue.

Chapter 9

A Computational Model to Detect Misunderstandings

This chapter presents an algorithm that detects misunderstandings in collaborative work at a distance. It analyses the movements of collaborators' eyes on the shared workspace, their utterances containing references about this workspace, and the availability of 'remote' deictic gestures. This method is based on two findings: 1. participants look at the locations they are talking about in their messages; 2. their gazes are more dense around these points compared to other random looks in the same timeframe. The algorithm associates the distance between the gazes of the emitter and gazes of the receiver of a message to the probability that the recipient did not understand the message.

9.1 Introduction

When collaborators are not co-located, their ability to work together is reduced. Chapter 3 discussed how during the last three decades, many solutions have been proposed to improve the efficiency of work at distance. The main philosophy of many of these was to increase bandwidth so as to better emulate a face-to-face interaction (Whittaker et al., 1993). As chapter 3 largely discussed, this approach is limited, because an increase of bandwidth is not always possible (Nardi, 2005), fruitful (Kraut et al., 2003), or advisable (e.g., air traffic control). One of the main conclusions of that discussion was to find valid alternatives for communication mechanisms that are effective in presence but not available in remote collaboration settings, as deictic gestures.

The same chapter described how pointing to an object in space leads the conversation participants to focus attention on that object, with a consequent disambiguation of context, and an economy of words used. Chapter 4 described many solutions that have been envisioned to overcome the lack of deixis at a distance. Early attempts consisted in *continuously* showing the mouse pointer of a collaborator on the partner's workspace (Gutwin & Greenberg, 2004), or through videoconference systems (M. Bauer et al., 1999). Recently, researchers proposed less invasive solutions allowing the user to choose when to display deixis to the partner (e.g., Mühlpfordt & Wessner, 2005).

The study reported in chapter 8 discussed how these sorts of virtual gestures might differ greatly from an actual movement of the finger over a map while face-to-face. In presence, deixis is always used in conjunction with eye gazing while this is not possible at distance, at least not with commonly available technology. The speaker who decides to use a deictic gesture can verify with his/her own gaze whether or not the listener has seen this gesture. In the same way, the listener also visually perceives communicative gestures. Gaze has, therefore, a double nature: it is both a perceptual and a communicative device.

Along this line, other solutions have been proposed to visualize and *continuously* share the position of the eyes of remote collaborators on a shared workspace (e.g., the GAZE groupware system developed by Vertegaal, 1999). More sophisticated solutions employed ad-hoc devices, like the ClearBoard system (Ishii & Kobayashi, 1992), or the gaze awareness display system proposed by Monk and Gale 2002). It is a shared conviction that eye-based interfaces offer enormous potential for efficient human-computer interaction, but also that the challenges for a proficient use of this technology lie in the difficulty of interpreting eye movements accurately. Just as it is difficult to infer comprehension from users' speech, eye-movement data analysis requires a fine mapping between the observed eye movements and the intentions of the user that produced them (Salvucci, 1999).

Pursuing this research direction, I argue that by combining multi-modalities of communication, such as deixis and gaze, we can build efficient solutions to support collaborative work at a distance without the burden of *continuously* displaying virtual finger or eye movements when they might not be relevant. An example of this approach is the RealTourist application proposed by Qvarfordt et al. (2005). To this end, the qualitative analysis that I reported in the previous chapter showed how participants of the 'festival' task, who produced utterances containing references on the shared map, often glanced at the spots they were indicating in their messages. Similarly, the recipients of these messages looked briefly at the points named in the content of the received messages. I hypothesized that this behavior was consistent in the different experimental conditions and that it could be used to detect episodes of miscomprehension between the remote collaborators.

In this chapter, I propose an algorithm which combines a language analysis of spatial expressions with gaze movements and virtual deictic gestures (see section 9.3). In order to build the proposed algorithm, some hypotheses needed to be verified, as discussed below. This analysis reported in this chapter has been conducted with the eye-tracking log files that I collected in the 'festival' experiment.

9.1.1 Hypotheses

The background discussion reported in chapter 3 suggests the hypothesis that indicating an object in space must also lead the participants to focus attention, and therefore gaze, on that object. This implies that effective indicating gestures should attract eye movements (Clark & Krych, 2004). Therefore, participant pairs communicating with tools supporting Explicit Referencing should better coordinate their eye-movements.

Richardson and Dale (2005) found that the higher the degree of eye movement coupling between a listener and a speaker, the better the listener comprehension. In chapter 8, I extended this finding to a synchronous and symmetrical task. These results in addition to the qualitative analysis reported in the previous chapter suggest that non-aligned eye-movements might be the due to the lack of comprehension between the collaborators.

This discussion highlights the importance of coordinating eye-movements for collaborative work (here called gaze convergence, with gaze divergence being its antonym), as being an indication of alignment of cognitive representation of the problem at hand. I thus produced the following hypotheses:

H1 When collaborators write, or read, utterances containing references to a shared map, they look at the points of the map they are talking about.

H2 Participant pairs using a tool supporting Explicit Referencing produce communication episodes with more convergent gazes.

H3 For a given utterance, gaze divergence on the shared map results in verbal repair acts.

H4 Pairs which account for higher gaze convergence should reach higher performance.

9.2 Methodology

This study extends the analysis of the experiment presented in the previous chapter. The data used to perform the analysis explained at section 9.3 was collected using eye-tracking displays during the festival experiment reported in chapter 7. Therefore, the method employed was already presented in section 7.2. As the conducted analysis required high quality eye-tracking data during the production and reading periods related to utterances containing references to the shared workspace, I had to exclude many studies that were not of suitable quality (as explained in section 9.4). Additionally, the participants pairs using ShoutSpace often moved the message window over the map, all the studies conducted with ShoutSpace could not be used for this analysis (the fixations on the message and those on the map could not be distinguished). I therefore decided to perform the study reported in this chapter only on one factor: the availability of Explicit Referencing mechanism, corresponding to the ConcertChat and MSN conditions.

9.2.1 Measures

I used the maximum score achieved during the 45 minutes as the measure of task performance. Most of the analyses presented in section 9.4 were performed at the utterance level, with the exception of the results in section 9.4.3, for which I averaged these episodic measures for each experiment. The variables that I studied were:

1. The distance between the gaze of the emitter on the shared workspace and the point(s) named in messages containing references to the map;

2. The distance between the gaze of the emitter of a message and the gaze of the recipient of the same message on the shared workspace;

3. The rate of misunderstanding: a dichotomous variable, showing the presence of repair acts in the 5 next messages following the message under examination containing a reference to the map;

9.3 Detecting misunderstanding

When remote collaborators need to coordinate actions on the shared workspace, they regulate the effort they put in their communication so to minimize the possibility of miscomprehension. However even with great care, misunderstandings are part of every human activity and an intrinsic characteristic of communication. Conversants are naturally accustomed to detect and repair miscomprehension, but the medium used to communicate influences this ability (Clark & Brennan, 1991).

At a distance, cues normally used by conversants to infer misreadings are not available. The aim of the algorithm presented in this work is to explore the possibility to offer computational support to compensate for this lack by detecting misunderstandings between collaborators. It uses the movements of collaborators' eyes on a shared plan and a linguistic model describing the content of their messages. It operates by associating the emitter's fixations on points of the map to those named in the produced message. Then, it outputs a score describing how closely the recipient has looked at the same points while the message was read and until his/her next action. Figure 9-1 describes the modules of this algorithm. Each important block is marked with a letter and will be detailed below.

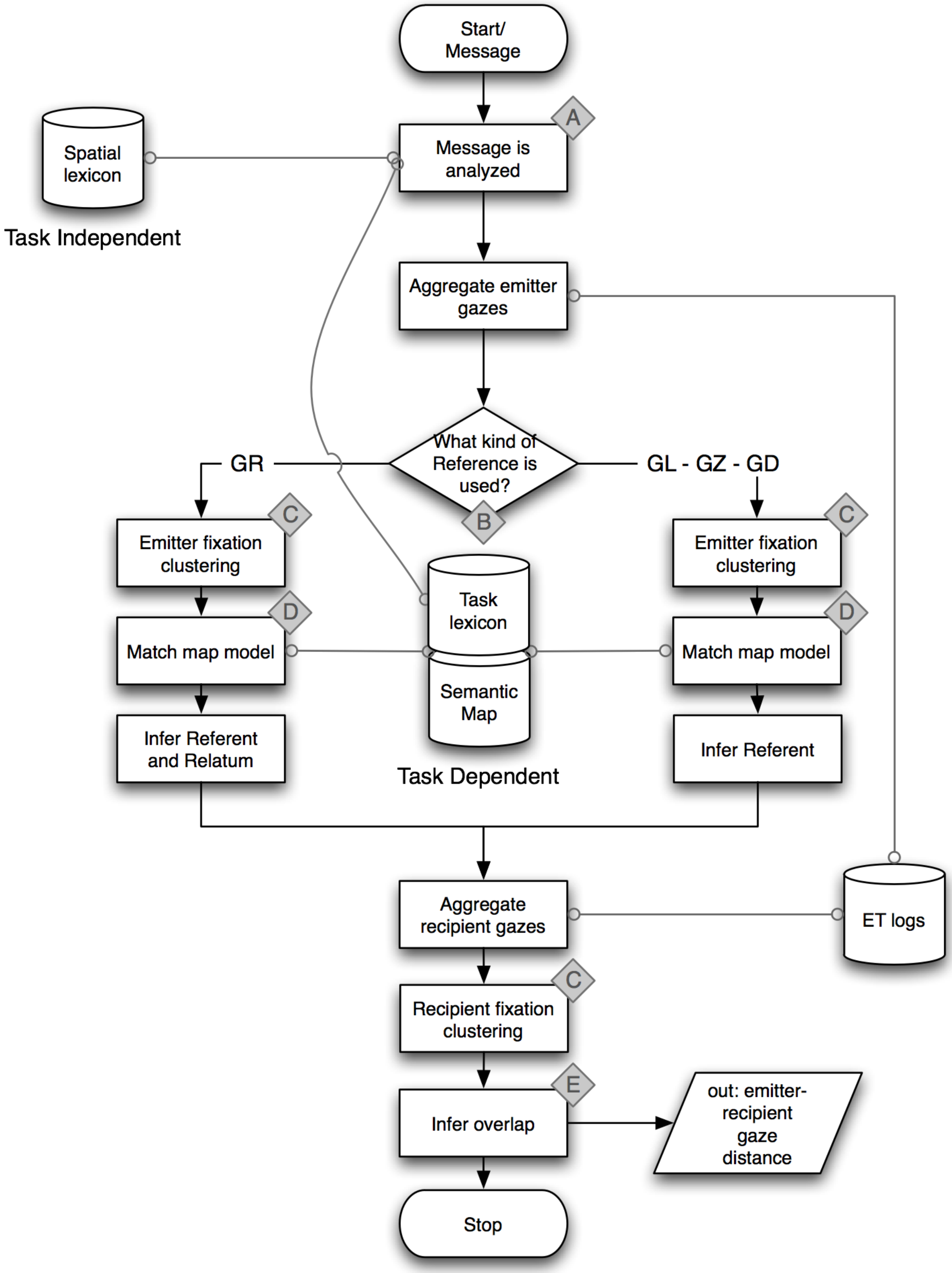


Figure 9-1: Algorithm proposed in this chapter to detect misunderstandings

9.3.1 (A) The linguistic model

Each message was parsed to detect and extract references to the shared workspace through a series of steps[1]. Two main class of elements were automatically extracted: the **Referent(s)** (the intended point(s) on the plan henceforth called the *interest* or *target polygons* , e.g., "*Parking 250*"), and the **Relatum(ii)** (the point used to identify the Referent, e.g., "*The green space on the right of parking 250*"), if any. When a relatum was present, I also extracted the **Relation** (the spatial correspondence between the Relatum and the Referent, e.g., "*on the right of*").

To accomplish this extraction process, typos were removed using a collection of common typing errors in French. Then the Relation was extracted by identifying prepositional phrases (e.g., "*on the right hand-side of the parking lot*", "*below the 'H'-shaped building*", etc.), and deictic expressions (e.g., "*I placed the second concert here*", or "*move your icon there*"). Matching was based on a task independent lexicon containing spatial referential expressions and it was adapted from Vandeloise (1991).

A second lexicon (task dependent) contained references to movable objects (e.g., 'the stage', 'the concert'). Finally, a third lexicon (also task dependent) contained references to fixed landmarks (e.g., "*P200*", "*chemistry building*"). These last expressions enriched my 'semantic map' of the university, a lexicon aggregating linguistic expressions pointing to specific polygons of my model of the campus (an excerpt is reported in appendix C, at page 339). These two task-dependent lexicons were used to extract the *interest polygons* used by the emitter in the block D described below (e.g., "*I want to use parking 200*" was associated to a polygon labeled P200).

9.3.2 (B) The Reference chooser

The detection of the three elements at block (A) triggered the assignment of the message to one of four possible categories of linguistic Relation (these are the same geo-reference strategies that were described in chapter 7):

1. A 'relative' positioning relations (GR) took advantage of visible landmarks to orientate the attention of the recipient towards the Referent (e.g., "*On the South of P250*");
2. A 'label' reference (GL) used visible or already established landmarks to orientate the recipient (e.g., "*let's use the P30*");
3. A 'zone' reference(GZ) was pointing to a portion of the map to reduce confusion (e.g., "*we can start on the upper-right corner*");
4. Finally, a 'deictic' reference (GD) used the Explicit Referencing mechanism offered by the interface or was used implicitly in the text to refer to landmarks established previously (e.g.,

[1] The parsing was performed by a Python script, which was supervised to correct mistakes and increase the quality of the descriptors of the parsed utterances.

"a stage here", or *"The place I told you before"*).

This scheme is quite simple but it is adapted to the specific needs of the 'festival' task, and it has the advantage that assignment to a category can be determined with non-overlapping rules such as: `[IF linguistic_deictic == True THEN category = GD]`. However, more complex categories might be used in different situations.

At this stage, I decided to differentiate between the kind of references used because in the case of utterances containing GR expressions, I could expect fixations either on the referent or the relatum, whereas in the other cases, fixations only occurred on the referent. This extra information was not used in the version of the algorithm tested in this work, however it might yield interesting applications in future refinements of the proposed mechanism. This point will be expanded in the discussion and in the conclusions of the thesis.

9.3.3 (C) Fixation clustering

This step consisted in aggregating the gaze data in order to infer zones attracting users' attention over a certain time period. The eye-trackers recorded the position of the eyes every 20 ms. A common way for aggregating such raw gaze data is to extract fixations which are defined by a very small position change between temporally continuous data. In my case, I needed a more general measure of the zone of interest over a rather long time (5-30 s.) compared to common fixation duration(<1s.) Thus, I developed a clustering method based on the density of raw gaze data disregarding the temporal order of the datapoints. First, the raw data were accumulated on a two-dimensional grid in order to compute a gaze density matrix. Then, the values contained in this matrix were smoothed with a gaussian filter[2] to eliminate discontinuities among its cells. Finally, a contour function was applied to this resulting density matrix. This resulted in a list of *isodensity* lines (see figure 9-2). Using the highest isolines of this list and a simple inclusion test to reject lower ones, I extracted a set of gaze-clusters. The centers of these gaze-clusters represented the *gaze density peaks*. More precisely, for each isoline taken in order of decreasing height, either it contained one of the already found peaks and then, it was added to it, or it was counted as a new peak if the peek number was not reached, in which case, the algorithm stopped.

I applied a simple optimization process to this clustering algorithm. The optimized parameters were the density grid size and the size and the variance of the gaussian filter. Two different evaluation-scores were used: the first one was the mean distance between the center of interest polygons (e.g., the parking lot named by the emitter of a message) and the gaze density peaks over all messages; the second evaluation-score was the number of messages in which the following was true: the computed gaze density peak, as defined above, was also the closest point to the reference

[2] In signal processing, it is a filter designed to give no overshoot to a step function input while maximising the rise and fall time. See `http://en.wikipedia.org/wiki/Gaussian_filter`, last retrieved March 2008.

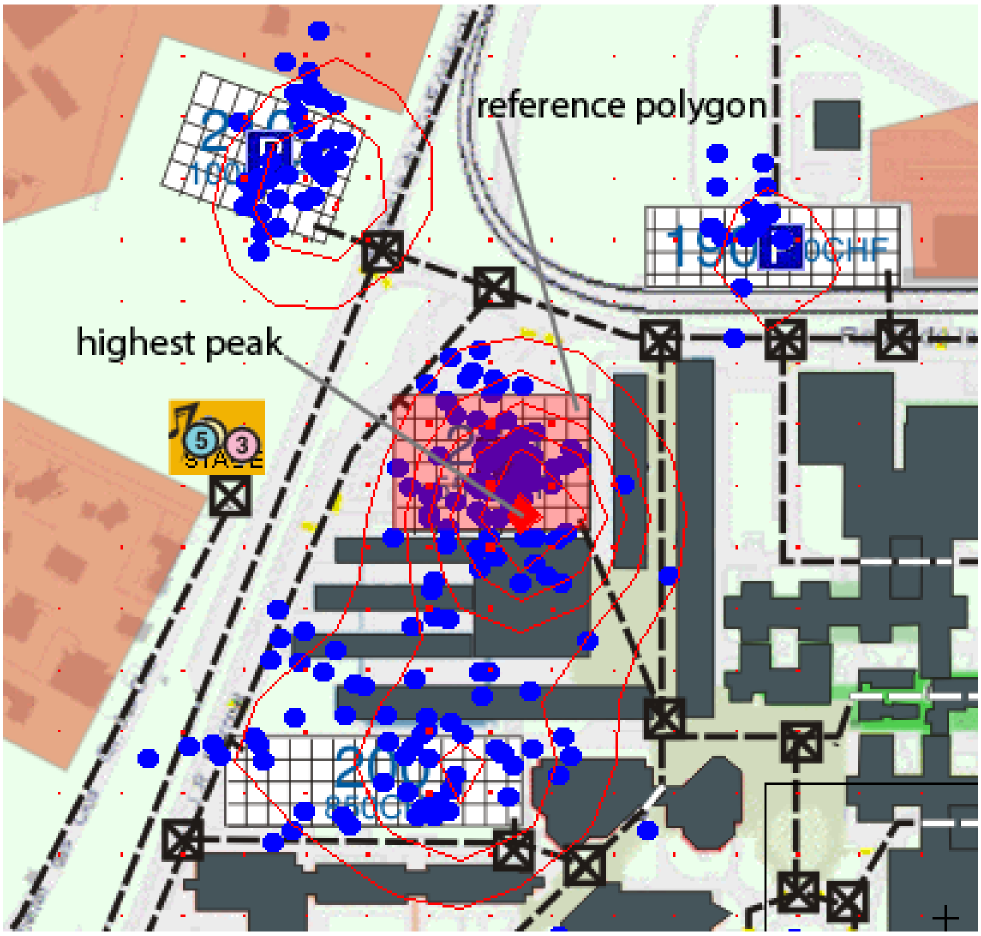


Figure 9-2: Example of contour plot used to calculate the clusters. The contour with the isoline corresponding to the highest density was associated to the interest polygon referenced in the message. Blue dots represent eye-tracker raw data

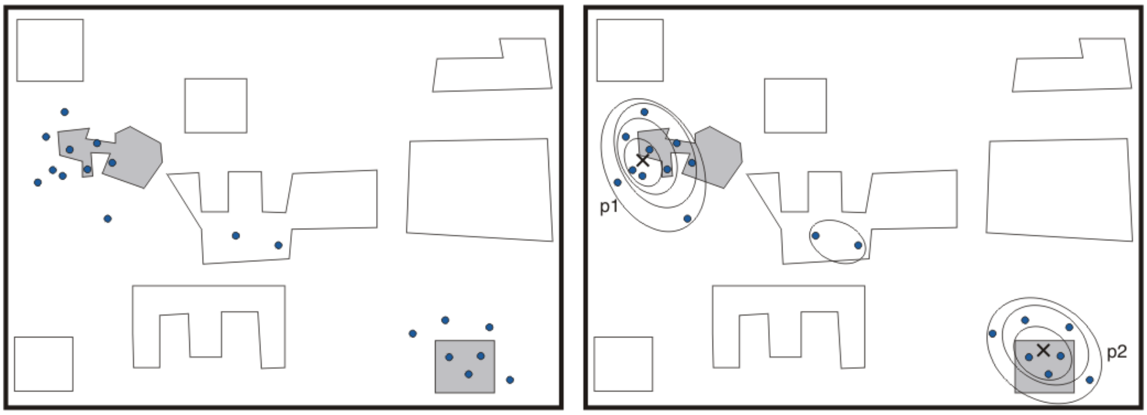


Figure 9-3: Map-matching process. (left) The gray-filled polygons are the target polygones, while the blue dots are the raw gaze movements over the map. (right) The gaze data is clustered in three groups. The peak of the top-left cluster (p1) is associated to the closest target polygon and the same goes for the bottom-right polygon. The third cluster in the center is discarded

polygon out all the centers of the other gaze-clusters eventually computed for a single message. These scores were computed for different combinations of the parameters used to compute the clusters: the size of the density grid and the size and variance of the gaussian filter.

The fitness landscapes for the optimized parameters showed some interactions between the parameters and moreover, the effects of the two evaluation scores were generally different and sometimes even opposed. For these reasons, it was difficult to find the best values for the parameters. I manually identified regions where one score was approximately stable and then I took the points associated with the best values on the other score for this selected region. The resulting optimized parameters were the following: density grid size: 15 by 11 squares (determines the amount of map surface that should be considered as part of the same cell), filter size: 6 squares and filter variance (σ): 1 (these are parameters of the gaussian filter applied over the grid for smoothing the values and might change with different maps).

9.3.4 (D) Match-map model

The resulting gaze density peaks had to be matched with the interest polygons extracted during linguistic analysis. This was mainly necessary to test the hypothesis that the writer (or the reader) of a message really looked at the referred objects. I used a very simple rule to find which peaks corresponded to which polygons when more than one peak were present. First, I simply took the N peaks with the highest isolines (where N is the number of interest polygons). Then, I computed the distances between each selected peak and each polygon. Finally, I associated every peak with the closest polygon, while checking that no polygon was associated to more than one peak (see figure 9-3).

In this analysis, the distance was simply measured between the center of the peak and the center of the polygon, but in a future version, it could be enhanced by computing the overlap between the peek contour shape and the polygon.

9.3.5 (E) Infer gaze overlap

In order to detect a possible misunderstanding between the two partners, a score recording the overlap between the writer's gaze density peaks and the reader's ones had to be computed. The only linguistic information used for this step was the number of interest zones (N): the N highest peaks of the writer were simply associated with the N highest reader peaks using the same methods than for associating peaks to polygons (see block D above) Then, the mean distance between the emitter's gaze density peak and the reader's gaze density peak associated with each interest polygon present in a message was taken as the miscomprehension score for that message.

In the future, this method could have been enhanced by taking into account the nature of the

Relation expressed in messages containing geo-references of the kind GR (see the last point of the discussion reported in section 9.5.1).

9.4 Results

Before testing my hypotheses I had to verify that the way I associated the gaze density peaks to the interest polygons extracted from the message was correct. As there might have been extra gaze density peaks calculated on the map, I used a simple matching rule: each interest polygon was associated with the closest gaze density peak (see figure 9-3). To validate whether this idea was correct, I used a binary vector of trials containing '1' when the associated gaze density peak was also that with the highest isoline available, otherwise '0' was used.

Contour-clustering of messages leading to a single gaze density peak were excluded from this analysis. A proportion test revealed that my assumption was correct 71% of the time (F[1,198]=66.27, p<.001), suggesting that the zones named in the messages were also those attracting the highest density of raw gaze.

9.4.1 Distance between contour-cluster and interest polygon

To test my initial hypothesis H1, namely whether the emitter and the receiver of a message containing a reference to the shared map were looking at the points named in the message (the *interest polygons*), I conducted an analysis of the distance of the emitter/receiver's gaze density peak to the interest polygon over time (more specifically, between -2 minutes and +2 minutes from the middle editing or reading time). Figure 9-5 shows the average distance between the gaze density peaks and the interest polygons at different time lags and according to the experimental condition (ConcertChat or MSN), while figure 9-4 shows this average distance at different time lags according to the phase during which a message was edited or read. I then subdivided this 4 minutes time interval in 99 segments that I used to sample the values of these two curves. These values were used to perform an analysis of variance with repeated measure that could determine whether the curves were statistically different.

The differences between the period of time where the message was edited or read and moments further or back in time were supported by a 2 (phase: editing–reading) × 2 (condition: MSN–cc) × 99 (lag times) mixed-effects analysis of variance (ANOVA) (lag and phase as a repeated measure factors) that revealed a significant effect of lag, F[98, 50060]=23.78, p<.001. By looking at figure 9-5, and figure 9-4, I could infer that this effect of lag is due to the central depression of the curve.

The interaction between lag and experimental condition was also significant, F[98, 50060]=60.75, p<.001, whereas the interaction of lag and phase was not significant, F[98, 50060]=0.37, p>0.1, n.s.. The same ANOVA revealed a significant effect of the availability of the Explicit Referencing

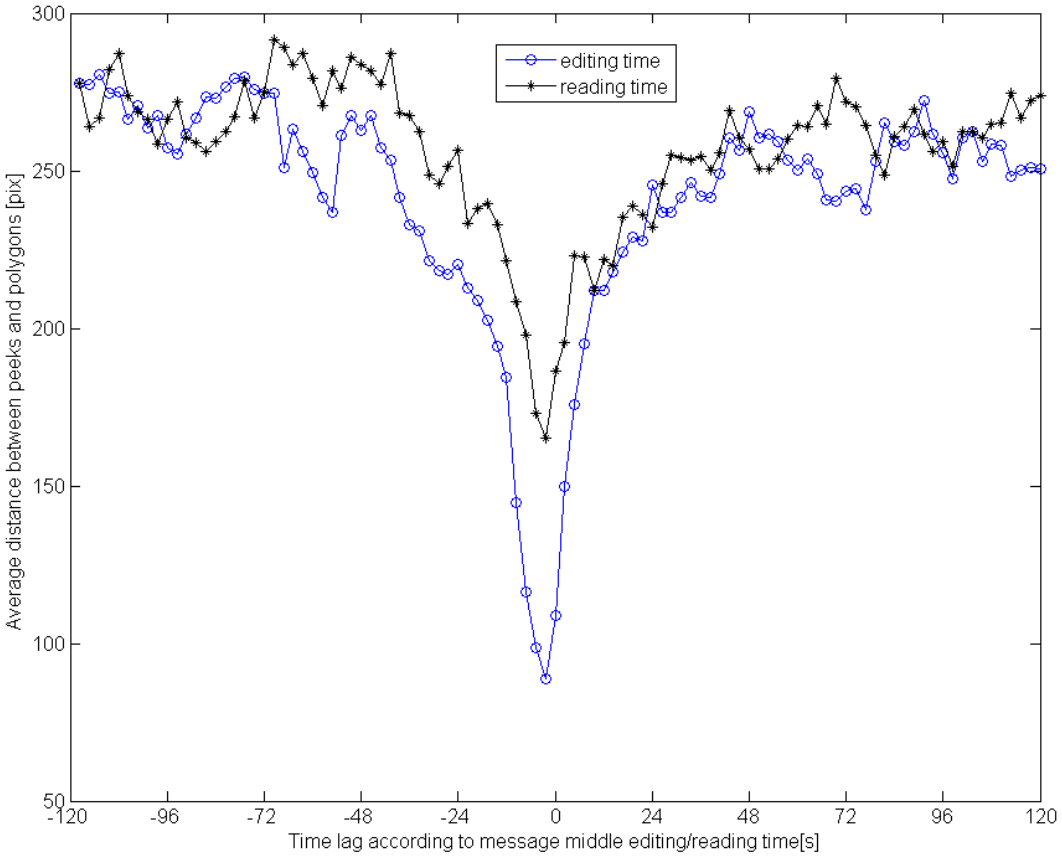


Figure 9-4: Relation between time and distance peek-polygon for the editing and reading phase

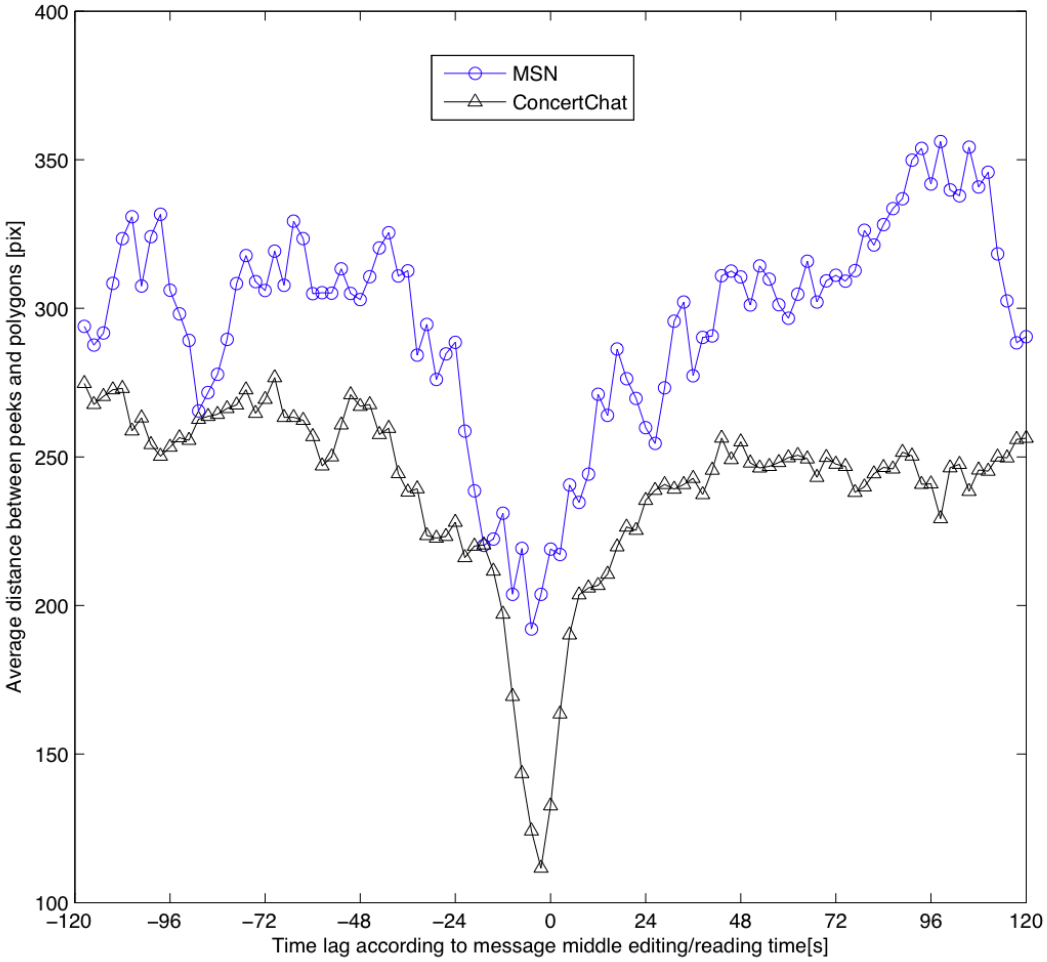


Figure 9-5: Relation between time and distance peek-polygon for two experimental conditions

mechanism, F[1, 374]=20.62, p<.001, and revealed a significant effect of the writing or reading phase, F[1, 50060]=23.78, p<.001.

I measured a mean distance in the reading phase of 263 pixels (*std* 107 pixels) *vs.* a mean distance in the editing phase of 252 pixels (*std* 107 pixels) for the period before -24 seconds and after +24 seconds. On the contrary, the minimum distance in the reading phase was 168 pixels (*std* 159 pixels) as opposed to 95 pixels in the editing phase (*std* 111 pixels) . This implies that, consistently with H1, emitters and recipients of a message containing spatial references to the shared map were looking at the interest polygon, during editing or reading, at above chance level.

Similarly, I measured a mean distance in the MSN condition of 291 pixels (*std* 128 pixels) *vs.* a mean distance in the ConcertChat condition of 251 pixels (*std* 101 pixels) for the period before -24 seconds and after +24 seconds. In contrast, the minimum distance in the MSN condition was 201 pixels (*std* 188 pixels) *vs.* 116 pixels (*std* 126 pixels) of the ConcertChat condition. If one consider the time before -24 seconds and after +24 seconds as a randomized level of eye-movements over the shared map, then it is possible to notice how the randomized gaze distance in the MSN condition (291 pixels) is bigger than the randomized gaze distance in the ConcertChat condition (251 pixels). This difference can be inputed to the different sizes of the maps in the interfaces of the two applications.

The difference in distance between the deepest points of the depressions of figure 9-5 is bigger than the difference between the extremes of the curves, thus validating the effect of the experimental condition (69.07% of the ratio of the extremes of the curves in the MSN condition *vs.* 46.21% of the same value in the ConcertChat condition). In other words, the difference between the curves around time '0' is bigger than before -24 and after +24 seconds. This last point is supported by the significant interaction effect between lag and condition which indicate that the central depression is deeper in the ConcertChat condition.

This finding was therefore consistent with H2, which predicted a smaller distance between the gaze of the emitter of a message and the gaze of the recipient in the trials conducted with the availability of tools implementing Explicit Referencing. Participants pair who used ConcertChat looked more closely to the target polygones during the task.

9.4.2 Writer-reader gaze distance and communicative repair acts

What happened when the distance between the gaze density peaks calculated for the writer and those of the reader was high? I will refer to this distance as *emitter-receiver peaks distance*. To answer my third hypothesis I assigned each message with a reference to the shared workspace to a dichotomous category. I looked at the five next messages following the message containing the geo-reference under consideration. If one of the replies of the reader expressed a doubt (e.g., "*Is it the rectangular building?*"), or asked for a repair act (e.g., "*I do not understand what you mean* *with building BM*"), in relation to the content of the original message then I marked the utterance as '1'. I marked '0' if that was not the case.

We performed a Kruskall-Wallis test which accounted for the difference of the distance between the emitter's gaze density peaks and those of the recipient of a message and the presence of a repair act in the following messages. Messages not followed by a repair act presented an average emitter-receiver peaks distance of 85.65 pixels, *vs.* messages followed by a repair act with an average emitter-receiver peaks distance of 231.37 pixels (χ^2[1,307]=206.03, p<.001). Figure 9-6 presents this relation graphically.

This finding is consistent with my third hypothesis, H3, which predicts a higher chance of a repair act after messages associated with divergent gazes between the emitter and the receiver. Therefore, when the emitter-receiver peaks distance exceeded about 100 pixels there was an higher chance of a possible misunderstanding between the participants than for shorter distances.

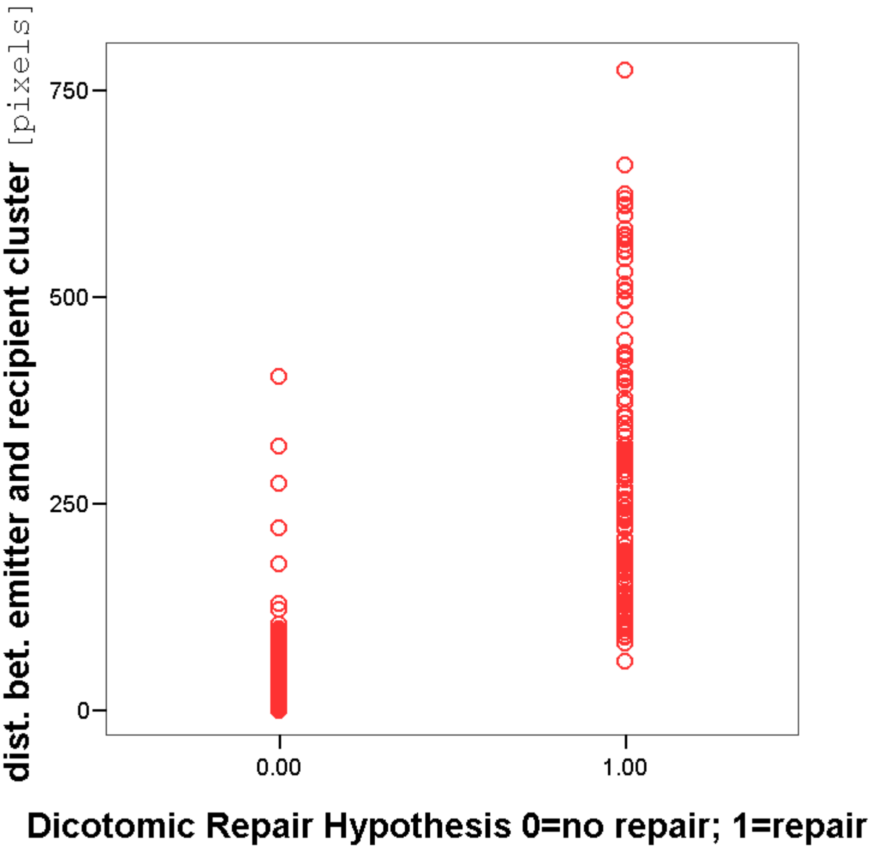


Figure 9-6: Relation between the distance of the peak of the emitter and that of the recipient and the dichotomous category coded to reveal the presence of a repair act

9.4.3 Relation of writer-reader gaze distance episodes with task performance

Contrary to my fourth hypothesis H4, I did not find a significative relation between the average distance of the emitter/receiver's gaze density peaks and the experiment performance (R^2_{adj}=–.01, β_{std}=–.21, p=.38, n.s.). Similarly, I did not find a significative relation between the number of communicative repair acts and the collaboration outcomes (R^2_{adj}=–.03, β_{std}=–.15, p>0.1, n.s.).

It must be noted that the number of episodes for which I managed to have a geo-reference on the shared map and valid eye-tracker data both for the writing and the reading phase was very low compared to the length of an experiment (45 minutes). I coded an average of ten messages per experiment, with an average duration of 30 seconds per utterance. Therefore, the part of the experiment covered by this analysis was roughly 11%.

9.5 Discussion

These study findings are consistent with the grounding theory (Clark, 1996). Results show that indicating and gazing are two interrelated mechanisms that are used to communicate. As predicted by Clark (2003), indicating an object in space must also lead the participants to focus attention on that object. The presence of indications, even if virtual as implemented by the Explicit Referencing mechanism, have an influence on the attention of the conversational partner.

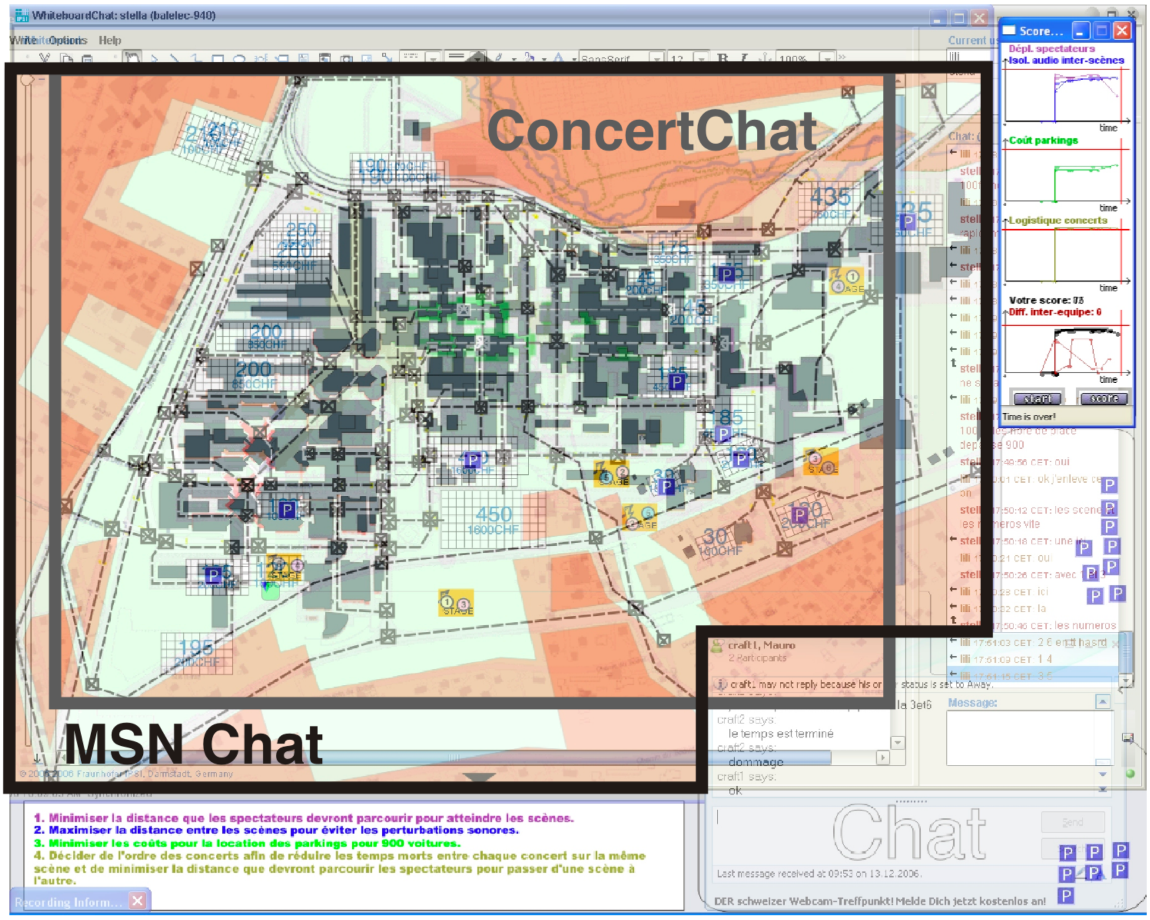
H1, When collaborators write, or read, utterances containing references to a shared map, they look at the points of the map they are talking about.

The findings presented here are also consistent with those of Clark and Krych (2004) showing that speakers monitor their own speech: a tendency was found to fixate on the polygons named in their messages at above chance level. This also extends the results of Griffin and Bock (2000) to written communication.

Similarly, the listener had the tendency to fixate on the part of the shared workspace named by the speaker at above chance level. This finding was consistent with those of Richardson and Dale (2005): I verified the importance of maintaining a consistent frame of reference during the interaction for the interlocutors.

H2, Participant pairs using a tool supporting Explicit Referencing produce communication episodes with more convergent gazes.

From the collaboration point of view, this study shows that the presence of an Explicit Referencing mechanism impacts the average distance of the collaborators' eye-movements. Observing the form of the depression of figure 9-5, I conclude that pairs using ConcertChat had more chances to be looking at the same map polygons when the emitter used a geo-reference. The differences between the remaining part of the curve of figure 9-5, which was considered as a randomized level of fixations, can be accounted for by the fact that the map in the ConcertChat condition was slightly smaller than in the MSN condition (see figure 9-7), thus resulting in a general shorter distance between random looks (the ratio of the average distance in ConcertChat and MSN in the extremes of the curves was 0.86, *vs.* the ratio of the sizes of the maps in the two conditions was 0.82). However, as explained in the results section, the difference of distance between the deepest points of the depressions of figure 9-5 is bigger than the difference between the extremes of the curves.



1. Minimiser la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. Maximiser la distance entre les scènes pour éviter les perturbations sonores.
3. Minimiser les coûts pour la location des parkings pour 900 voitures.
4. Décider de l'ordre des concerts afin de réduire les temps morts entre chaque concert sur la même scène et de minimiser la distance que devront parcourir les spectateurs pour passer d'une scène à l'autre.

craft1, Mauro
2 Participants
craft1 may not reply because his or her status is set to Away.
craft2 says:
le temps est terminé
craft2 says:
dommage
craft1 says:
ok
Last message received at 09:53 on 13.12.2006.
DER schweizer Webcam-Treffpunkt! Melde Dich jetzt kostenlos an!

lilli 17:51:03 CET: 2 6 en tt hasrd
lilli 17:51:09 CET: 1 4
lilli 17:51:15 CET: 3 5

Figure 9-7: Comparison of the size of the map in the ConcertChat condition (gray) and in the MSN condition (black). The ratio of the sizes of the maps in the two conditions was 0.82

H3, For a given utterance, gaze divergence on the shared map results in verbal repair acts.

This study also shows that a shorter distance between the fixations of the collaborators is related with fewer misunderstandings: the shorter the distance between the emitter-receiver peaks distance, the smaller the chance that the recipient of the message will express a doubt or will ask for a repair act. The importance of repair acts to the collaborative process was highlighted by Clark (1996). It must be noted that misunderstandings do not necessarily impact negatively collaboration. In collaborative learning situations, for instance, the presence of misunderstandings can foster deeper explanations and meta-cognitive processes that are positively related with the collaboration outcomes (Dillenbourg, Traum, & Schneider, 1996; Weiss & Dillenbourg, 1999).

H4, Pairs which account for higher gaze convergence should reach higher performance.
Further work is required to draw conclusions about the relation of the emitter-receiver peaks distance and the task performance as the analysis that was conducted referred to micro-episodes of collaboration with a short length. Taken their duration all together, these segments represents only a tenth of the task time and therefore their variability cannot account for the complete collaboration process. This small number of messages coded was due to the low quality of the eye-tracking data and the particular kind of utterances that were selected. In fact, participants often assumed bad postures during the 45 minutes of the task, creating 'holes' in the eye-tracker data. To validate the proposed algorithm, only messages containing references to fixed landmarks of the map and with valid eye-tracker datapoints both for the writing phase and the reading phase were considered.

9.5.1 Possible improvements of the algorithm

The test of this algorithm necessitated a detailed semantic map of my university campus (the task dependent lexicons described in section 9.3.1; see also an excerpt in appendix C, at page **??**). This map contained a precise definition of all the polygons/shapes that could have been recognized as functional for the task (e.g., buildings, road, crossings, parking lots, etc.). Each polygons was associated with a list of names that were usually used in the chat conversations to refer to it. This information allowed me to test the second hypothesis presented in this paper. Given the acceptance of this hypothesis for granted, further implementations of this algorithm might use a simpler linguistic model. Knowing how many objects are named in the utterance might suffice for the selection of the relevant cluster(s) in the rest of the procedure. This point will be developed in the final chapter of the thesis.

Finally, as suggested in section 9.3.4 the way peaks are associated to polygones can be further improved by taking into account the surface of the polygon being overlapped by the iso-density lines of the gaze cluster.

9.6 Conclusions

This chapter presents an algorithm which combines a linguistic model with eye-tracking data and that was proven to be effective in detecting misunderstandings in the experimental task presented in the previous chapters. The implemented linguistic parser had the basic function of inferring the number of zones that could have been potentially implicated in the communication exchange. This information was then used by the clustering module to associate the contours containing the highest peaks to their reference landmarks, and to discard not-relevant extra fixations on the map. This mechanism was meaningful because participants tended to look at the objects they were talking about beyond the level of chance, and because their raw eye-movements were more dense close to the polygons named in the exchanged messages. These basic findings allowed the design of the algorithm presented here that can be used in real-time applications to detect and eventually signal misunderstandings.

The core of this algorithm is the combination of modalities of communication, namely the movements of the collaborators' eyes on the shared workspace, and their utterances possibly containing references on this plan.

Further work is required to verify the findings presented here: testing the algorithm in real-time; using a different metrics to calculate the misalignment of the collaborators (e.g., using an area of overlap between the collaborators' gaze clusters instead of the mere distance of their centers); using a different task, with different communication tools.

This work aims at showing the potentials of using a simple interaction mechanism to sustain collaborative work at a distance. My main argument is that it is not necessary to implement interfaces aimed at imitating face-to-face with a continuous increase of bandwidth. Instead, simpler solutions, like the combination of gaze and remote deixis presented here, offer greater potential.

Chapter 10

General discussion

This final chapter summarizes the results, limits and contributions of this thesis and describes the potential implications for the design of applications supporting collaborative work at distance. Additionally, this chapter will indicate future perspectives on this work.

10.1 Summary of the contributions

This thesis has explored the use of Collaborative Annotation Systems to support remote collaboration. The thesis explored the idea that shared annotations of maps can be used to support remote collaboration in tasks requiring spatial coordination.

In chapter 2, I discussed relevant literature on the interconnections of language, space and human interaction in the general context of this thesis. In particular, I focused on the mutual relation of space and language: given a certain scene containing objects arranged in a certain spatial configuration, only a limited number of descriptions can communicate the arrangement of the objects contained in the scene. Conversely, given one such spatial descriptions, only a limited number of interpretations are possible.

Spatial descriptions form cognitive maps in people's minds and these influence how people think and behave. Interpretations of descriptions are limited because language is understood through a combination of linguistic and extra-linguistic constraints. These last involve our knowledge of the world, but also information coming from multiple communication channels, like deictic gestures, which connect the space in which the communicative interaction occurs and the linguistic expressions which are produced by the conversants.

Chapter 3 described the various lines of research that are important in framing the research questions and experiments of this thesis. In particular, the chapter focused on deictic gestures explaining how this mechanism is a basic pragmatic need of communication. This same chapter gave examples of technological solutions that aimed at supporting collaborative work at a distance but failed in this attempt because they fractured the ecology of the interaction they were trying to sustain. This failure was attributed to what is described as *imitation bias* (Dillenbourg, 2003): the attempt to replicate the same modalities of face-to-face communication for supporting remote collaboration. There are reasons to believe that the more we try to quantitatively improve the connection (e.g., the bandwidth) between remote sites, the more we might exacerbate the fracture of their ecologies. The chapter described relevant literature to show how communication is a cooperative process, implying that a maximally shared visual environment is not required to obtain the best comprehension efficiency. The solution is to find valid alternatives for mechanisms which are efficient while face-to-face but not available when at distance.

This thesis focuses on deixis, and in particular deictic gestures. This is the ability to employ multiple modalities of communication to disambiguate content and to reduce the effort required to reach a common understanding. A deictic gesture binds a particular utterance to a point within the sight of interlocutors. This mechanism is naturally embodied in conversation when people are face-to-face, but it is not available when conversants are not co-located. However, different technological solutions are possible to enable Explicit Referencing. In particular, the works presented chapter 3 suggest the hypothesis that different designs of Explicit Referencing mechanisms can have an impact on collaboration. Particularly, the criteria under which messages are organized might structure the conversation following a temporal or spatial order. Finally, one last important feature of deictic gestures was highlighted: when in presence, these acts are naturally perceived and acknowledged through gaze.

Chapter 4 presented a non-exhaustive list of Collaborative Annotation Systems, or CAS. These are applications that implement, in different ways, the mechanism of Explicit Referencing. Many of the presented applications specifically targeted the annotation of shared maps. This chapter presented my first theoretical contribution: a **taxonomy of CAS**, which describes these tools according to eight dimensions. In particular, I chose to focus the rest of the work on three features of CAS: *the degree of immersion* for which the tool was designed, this being either for mobile or fixed use; *the presentation of the referencing link*, being overlaid on, or placed to the side of the shared workspace. This last dimension was connected to *the organization criterion of the messages*: messages could be retrieved following a time criterion or they could be displayed by the context to which they referred.

Chapter 6 reported the qualitative observations conducted with the ubiquitous CAS named STAMPS. Using this application, a field trial was conducted in Geneva, involving participants from different communities and age groups. My objective was to understand *how* and *why* people would associate virtual messages to physical locations and which dimensions should be considered relevant to the design of an application supporting this activity. The approach followed to answer these questions was to monitor and record the actions of the users in the system and to analyze their traces looking for patterns of use among participants. The trial lasted three months. A low level of participation was recorded. This was attributed to a lack of useful content in the system, and a lack of a critical mass of users.

The second lesson that can be taken from this thesis is that **the criterion under which the messages are organized in the interface has an influence on the way users produce and retrieve content with a CAS tool**. Messages in STAMPS were organized by the spatial content to which they were attached. This choice influenced the users to mainly retrieve content by browsing the map. This subsequently contributed to the production of asynchronous communications by users. In follow-up interviews, participants clearly stated that it was difficult to keep track of conversations. Therefore, I chose to test quantitatively, in a controlled experiment, how the organization of the messages could influence collaboration.

A third contribution of this work stemmed from the second experiment, a field trial organized with students of urban planning in Lausanne. The study demonstrated how **under specific task requirements, the user of a CAS tool would interact differently from the way participants of the first trial interacted** with the tool. The analysis of the logs showed how communication was mainly self-directed. In particular, annotations were designed and deployed by the note-taker herself.

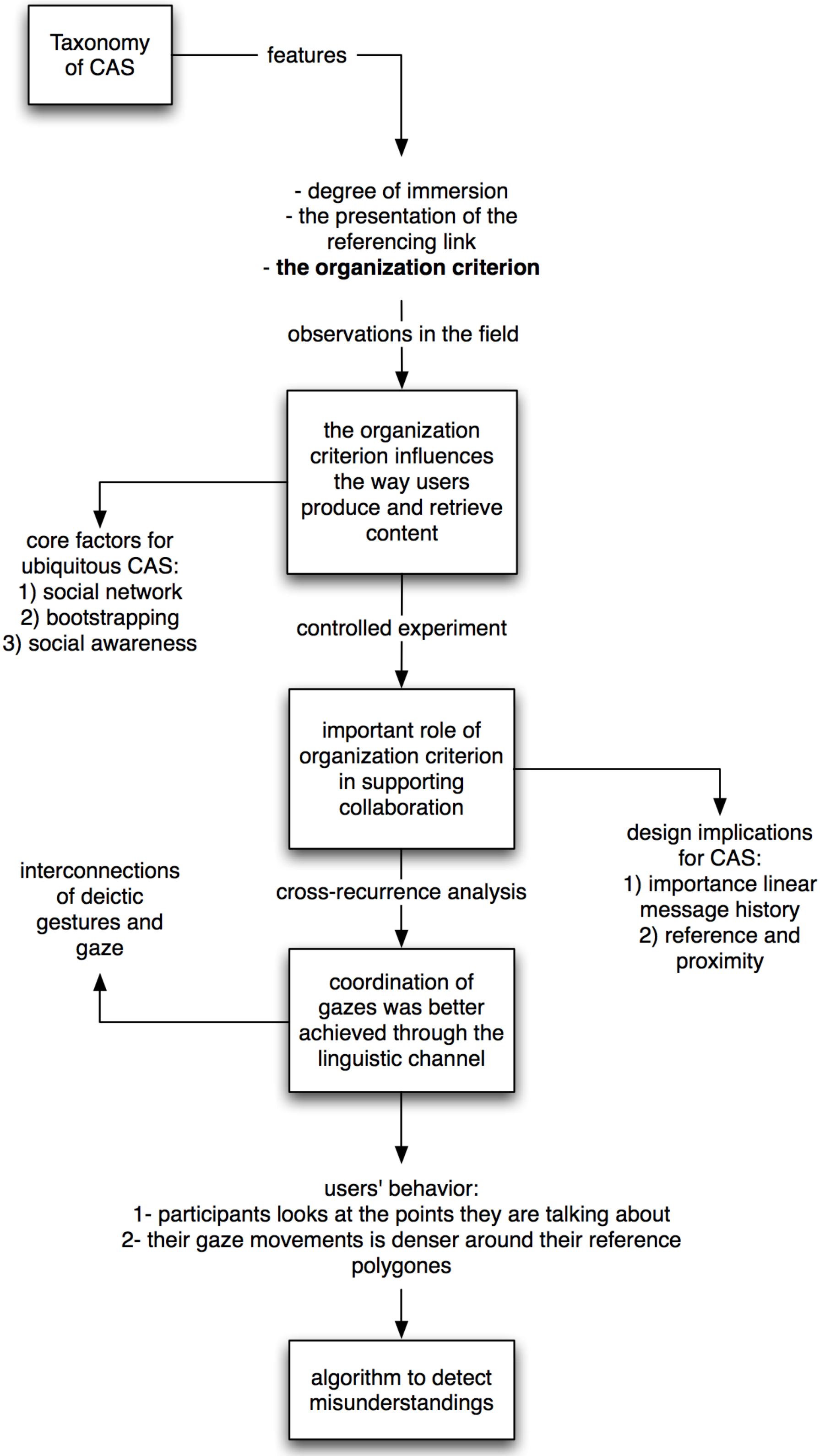
The main results of the quantitative experiment of this thesis are reported in chapter 7, where I presented a quasi-experimental design to compare the influence of applications offering a different degree of support to Explicit Referencing to remote collaborations. In this experiment, I varied the availability of an Explicit Referencing mechanism and the presence of a linear message history (corresponding to the organization criterion discussed above) in the communication tool used by the participants to coordinate their efforts. This chapter reported the fourth contribution of this thesis, in demonstrating **the more important role of a linear message history to the presence of an Explicit Referencing mechanism in supporting collaboration at a distance**. This study demonstrated the adaptability of human communication to different media configurations. Participants across the different conditions adapted their communication styles to the different tools they were using. Participants with a linear message history wrote many short messages with a 'messy' turn-taking process, while participants in the opposite condition wrote fewer longer messages with a more regular turn-taking to avoid confusion. Pairs using a tool implementing an Explicit Referencing mechanism took full advantage of the deixis offered by their tool and, whenever needed, created references to the shared workspace. All these results are reflected by the mean editing time of the messages exchanged by a pair (see section 7.3.5, at page 207). This variable aggregates the effects of many other process variables and mediates the manipulation of the independent variables on the dependent variables.

Chapter 8 looked at the interaction of Explicit Referencing and how participants looked at the shared workspace. The results of the 'festival' experiment were analyzed with a cross-recurrence analysis to compare the two time-series of the eye movements. The analysis did not reveal an influence of the experimental conditions on the amount of gaze coupling. However, a correlation between the amount of gaze coupling and the task performance was found. Also, a relation between the frequency of deixis expressed through the linguistic channel (anaphora) and the amount of gaze coupling of a pair was found. The fifth contribution of this thesis lies in the interpretation of this result: the **coordination of gazes was better achieved through the linguistic channel than through the visual channel supported by the Explicit Referencing mechanism**. More research is required to validate these findings and to understand whether they were caused by the lack of visual acknowledgement of the messages containing a reference to the shared workspace. Nevertheless, this chapter suggests the existence of a deep interrelation between the mechanisms of deictic gestures and gaze. Their dissociation, which may occur in the design of CAS, negatively impacts collaboration.

The same chapter, included a qualitative analysis of the eye-movements during production and retrieval of sentences. The analysis revealed a recurring behavior of the emitter of a message: whenever she was producing a message containing a reference to the shared workspace, she was glancing at the polygon named in the message right before or while composing the text. Similarly, the recipient of the message would glance at the reference polygon while reading the message or right after. These findings suggested the idea of building a computational support for this activity that could take into account the distance between the points looked at by the emitter and the recipient and the references contained in the text in order to account for occurring misunderstandings between collaborators.

The hypothesis suggested by these last findings were further analysed in chapter 9, where an algorithm for detecting automatically miscomprehension episodes between participants solving the 'festival' task was proposed. The design of the algorithm relied on two confirmed hypotheses: (1) participants looked at the points they were talking about in their messages; and (2) during the production and the retrieval of the message, the eye movements of the emitter and the recipient, respectively, were denser around these points compared to any other region looked at in the same period. The algorithm extracted the number of references used in the exchanged messages and used this information to cluster the raw movements of the eyes of the participants on the shared workspace. Then, it associated the distance of the denser gazes of the emitter and the receiver of a message with the probability that the recipient did not understand the emitter's message.

The algorithm was tested against the eye-tracking logs collected during the 'festival' experiment. Due to quality of these logs, only a limited number of messages could have been treated, accounting for only 10% of the task-time. Each parsed message was analyzed, looking for the



Taxonomy of CAS
features
- degree of immersion
- the presentation of the referencing link
- **the organization criterion**
observations in the field
the organization criterion influences the way users produce and retrieve content
core factors for ubiquitous CAS:
1) social network
2) bootstrapping
3) social awareness
controlled experiment
important role of organization criterion in supporting collaboration
design implications for CAS:
1) importance linear message history
2) reference and proximity
interconnections of deictic gestures and gaze
cross-recurrence analysis
coordination of gazes was better achieved through the linguistic channel
users' behavior:
1- participants looks at the points they are talking about
2- their gaze movements is denser around their reference polygones
algorithm to detect misunderstandings

Figure 10-1: Summary of the contributions of this thesis

presence of linguistic repairs in the conversation that followed. A factor that was coded in a dichotomous variable. The analysis of this result revealed that the distance in pixels between the emitter and the receiver *gaze density peaks*, as defined in chapter 9, was highly related to the presence of a repair act in the conversation that followed a message. This constituted the sixth contribution of this thesis, as it demonstrated **the ability of the algorithm to capture episodes of misunderstanding on the basis of gaze patterns**.

Figure 10-1 summarizes this discussion.

10.2 Limits of the studies presented in this thesis

The three experiments presented in this thesis explore how different designs of CAS systems can impact collaboration. However, these studies are subject to limitations and therefore, their generalization to other settings or populations must be considered in light of these shortcomings.

The first qualitative study was limited by many factors that emerged during the early weeks of the field trial. Mainly, the participation of the users was limited by the fact that STAMPS was not available to the participants' social networks. As interviewees clearly stated, the production of content lacked the incentives that are provided by other social platforms. Participants producing content in other systems are encouraged because their social image benefits from friends accessing their content. During the field trial, users did not receive these sorts of social incentives because the group-mates of each participant were not her closest friends. A second factor that negatively influenced the use of STAMPS was the lack of useful content in the system. Although, I tried to bootstrap STAMPS with touristic information, this was not helpful to the participants. These two choices now appear critical and should be reconsidered carefully in further trials.

The design of the first and second qualitative experiments should also be reconsidered carefully. First, the length of the trials was too long. People were extremely excited to test a new technology in the initial weeks. However, this initial enthusiasm was reduced by the usability problems of the application and the passing time. The particular timeframe of deployment of the first experiment was also not optimal as the study was conducted over the summer and many students had to leave for their holiday locations. A different choice would have been to conduct a series of shorter field trial lasting two/three weeks and interwoven by reflection periods during which the application could have been improved. Finally, the choice of not offering a structured scenario of use might have been critical to the obtained results. A different option would have been to test several scenarios and the conditions under which they could positively influence the production of messages.

This is also connected with another major limitation of the qualitative study: the lack of a pre-study revealing positive conditions for deployment of a collaborative annotation system. I did not run an initial observation to understand whether alternative practices were already in place among the groups of participants and how STAMPS could have replaced or complemented existing social rituals or practices.

The quantitative study was also limited by several factors. First, the reported results were obtained from a single experiment designed on the task of organising a music festival. For solving this particular problem, participants had to rely strongly on the use of spatial descriptions to achieve successful coordination. However, it would have been interesting to manipulate this need for using spatial references using a different task, and observe whether this could have had an effect on performance across experimental conditions. For instance, it would be interesting to see whether the same results hold in an experimental condition reintroducing the *feedthrough* that I intentionally removed (e.g., in the specific case of the 'festival' experiment, the movements of the icons would be synchronized across displays).

Additionally, the results of the quantitative experiment were limited by small differences in the features of the chat tool used. For instance ConcertChat and MSN chat offered an "awareness" mechanism that allowed the users to know when their partner was composing a message. ShoutSpace offered a threading functionality between the messages that allowed the participants to follow conversations about the same topics (part (w) of figure 7-3, at page 189). This was not possible in the other conditions. Another difference between the interfaces was that the number of messages in the history of ConcertChat was slightly higher than that of MSN chat. However, when analyzing the logs of eye-trackers, I saw that participants rarely made fixations on messages older than the fifth previous one. Finally, the surface of the map in the ConcertChat condition was slightly smaller than that on the other three conditions (see figure 9-7, at page 260). This difference was compensated by the mean of adapting the way a statistic measure was computed or by testing whether the found effect was due to the different map sizes. For instance in chapter 8, this difference was compensated by reducing the threshold parameter used to calculate the recurrence (as explained in section 8.4.1, at page 230). Conversely in chapter 9, I tested specifically whether the differences in the obtained results were generated by this different map size. The analysis refuted this hypothesis, as explained in section 9.5, at page 259. The reason why I decided to employ communication tools that offered an unequal set of features was to maintaining a minimal ecological validity in the experiment (Brewer, 2000; Shadish et al., 2002). In fact, all the participants reported to have previous communication experiences with Microsoft MSN chat.

Another important point is that the methodology of cross-recurrence presented in chapter 8 and the clustering solution derived in chapter 9 should be tested in different domains to validate their effectiveness.

To conclude, it would be interesting and useful to reiterate this experiment with a different task, with a different number of participants (e.g., 3 to 6), and with more communication tools that are more homogeneous in terms of the interface features that are not subject to experimental manipulation.

10.3 The reasons for failure of a location-based annotation service

The research with STAMPS enabled me to reflect on the reasons why, although acclaimed for years, location-based annotation services are yet to become a common communication mechanism. First and foremost, each new service has an **adoption cost** for the user. Field trial participants were asked to use a different phone from their own in order to be able to participate. This limited adoption as people might be accustomed to a particular mobile for reasons other than technical specifications (e.g., 'look and feel'). The value gained from the use of the service must exceed the adoption cost, this was not the case for STAMPS.

These services are not used in mainstream communication because they never reached **critical social mass**. This thesis shows how the production effort of a prototypical user of CAS systems might be rewarded by the exposure of her content to her group of peers. The success of these systems results from the tradeoff between the expenditure of leisure time and the personal benefit that an author of content might receive from her friends accessing it. The reasons why these messages are not accessible, which explain why this critical mass was not reached is related to the lack of standards discussed in point four below.

A second factor that emerged that could be responsible for the missing adoption is the **lack of awareness or lack of certainty about the interaction mechanism** of the service. For instance, applications might reveal inconsistent behavior during manipulation or might not offer appropriate feedback for executed actions. People might feel uncertain that the application will assist in accomplishing their communication intentions. In some particular occasions, the delivery of a message under precise circumstances is vital to the usefulness of the application, and therefore to its adoption.

A third factor are the **ergonomic barriers** to which a service might expose its users. Usability issues for applications that require extremely complicate installation procedures are examples of this category of failure. This was not the case for STAMPS as it came already installed on users' phones. However, during the field trial I released some upgrades and fixes that were installed by few participants for the reasons expressed in this paragraph. Also, the design of the interaction mechanism with which users are expected to use the service might bias the production and the retrieval of content in the system. Therefore, designers should carefully consider how the organization of content in the system and other factors related to the features of its interface might affect the number of scenarios of use that will be supported by the designed mechanisms and those that the users will choose to employ in their daily routines. In short, “details kill”!

One final limiting factor that the field trial highlighted is the **lack of standards**. Devices offer inconsistent features and APIs[1]. In particular there are no widespread mobile platforms[2]. Each mobile manufacturer produces its own software developer kit which makes mobile software development extremely costly. While developing STAMPS, I directly experienced how programming for mobile phones is extremely complicated and time consuming.

10.4 Design implications

Although limited by the weakness described in section 10.2, the experiments presented in this thesis can yield relevant implications for the design of location-awareness tools for collaborative tasks. The principles described in this section remain highly theoretical and they are not meant to offer practical recipes.

10.4.1 Annotating maps for supporting collaboration and communication: the importance of a linear conversation

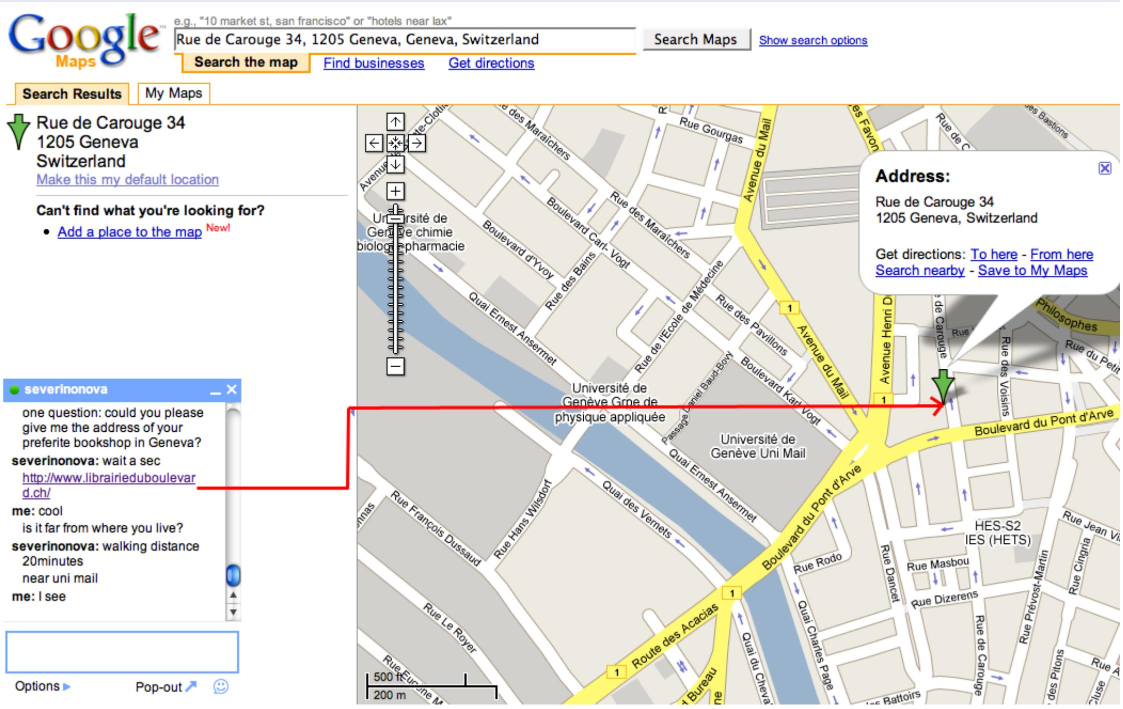
The results presented in chapter 7 support the idea that the linearity of conversation is more important than the presence of an Explicit Referencing mechanisms in the design of interfaces aimed at supporting collaborative work at a distance. The experimental manipulation of a linear message history influenced many of the variables analyzed for the ‘festival’ experiment. The presence of a linear message history also influenced performance. In particular, participants using a tool implementing a linear message history reached the placement of the icons of the task faster than participants in the opposite condition. Additionally, pairs with a linear message history tested a higher number of different solutions than pairs that did not have a linear message history.

It is reasonable to extend these results to other applications and domains claiming that the presence of a linear message history allows participants to: a) **decrease their cognitive load** by externalizing static information to the most persistent medium. For instance, in the festival experiment, the order of the concert could remain visible and therefore accessible through the message history, thus facilitating collaboration; b) **enable parallel contributions** to the task. For instance, in the situation of a collaborator (‘the leader’) giving directions to the ‘the follower’ on how to place a number of icons, the leader can prepare the next instruction to be given, while the follower moves one piece. This economise task resolution time. Gergle et al. (2004)

[1] An application programming interface (API) is a source code interface that an operating system, library or service provides to support requests made by computer programs.

[2] STAMPS was developed for the Symbian platform. At that time, this was the only solution that allowed to acces low-level functionalities of the phone, like the antenna identifier. Recently other companies are developing alternative solutions which might offer other possibilities (e.g., Google with the Android platform and Apple with the iPhone platform).

demonstrated many of the properties of a linear message history in supporting collaborative work at a distance. The results presented in this thesis extend their findings by explaining that the role of a linear message history was more important than that of Explicit Referencing in upholding remote collaboration.



e.g., "10 market st, san francisco" or "hotels near lax"

Rue de Carouge 34, 1205 Geneva, Geneva, Switzerland

Search the map | Find businesses | Get directions

Search Results | My Maps

Rue de Carouge 34
1205 Geneva
Switzerland
Make this my default location

Can't find what you're looking for?
- Add a place to the map New!

severinonova

one question: could you please give me the address of your preferite bookshop in Geneva?
severinonova: wait a sec
http://www.librairieduboulevard.ch/
me: cool
is it far from where you live?
severinonova: walking distance
20minutes
near uni mail
me: I see

Address:
Rue de Carouge 34
1205 Geneva, Switzerland

Get directions: To here - From here
Search nearby - Save to My Maps

Figure 10-2: A possible way to improve the mapping services of GoogleMaps to support synchronous interactions. At the moment it is not possible to have chat conversation in front of a map. Also the current chat application does not allow for drawing on the workspace

Designers of CAS, should carefully consider how to allocate the interface real-estate to support conversation and referencing to the shared workspace. For instance, the current design of GoogleMaps is not meant to support synchronous interaction. Users can produce geo-references but there is no direct support for sharing this content in a real-time conversation. A sticky-notes approach, as used in the current design of the interface, affords asynchronous usage scenarios. Figure 10-2 presents a possible improvement of the interface of GoogleMaps so as to add support for synchronous collaboration. The current mapping service does not allow chat conversations while looking at the same map. Also, the chat application that is available in the GoogleDocs[3] service does not allow for the drawing of lines on the workspace and therefore it does not support Explicit Referencing to a shared document or map. These limitations could be technically solved allowing synchronous collaborations over the map service or other shared documents (web page, text or spreadsheet).

[3] Online collaborative editing service. See `http://docs.google.com/`, last retrieved March 2008.

10.4.2 Reference does not require proximity

Another way to look at the results presented in this thesis is that reference does not require proximity. For instance, a message `M` may refer to point `X` with a deictic `M`—`X` link, with positive cognitive implications even if the message is not displayed at the location `X`. A *visual link* between the two might be just sufficient to communicate the reference to the shared workspace. By visual link, I mean an arrow or other marker that can visually and continuously connect the point where the message is displayed and the reference point on the workspace. Symbolic links are numbers or symbolic icons which are overlaid on the map and duplicated close to the messages that relate to the map. Symbolic links might not be as efficient as visual links. They require an extra effort to mentally match a symbol to its anchor point on the map.

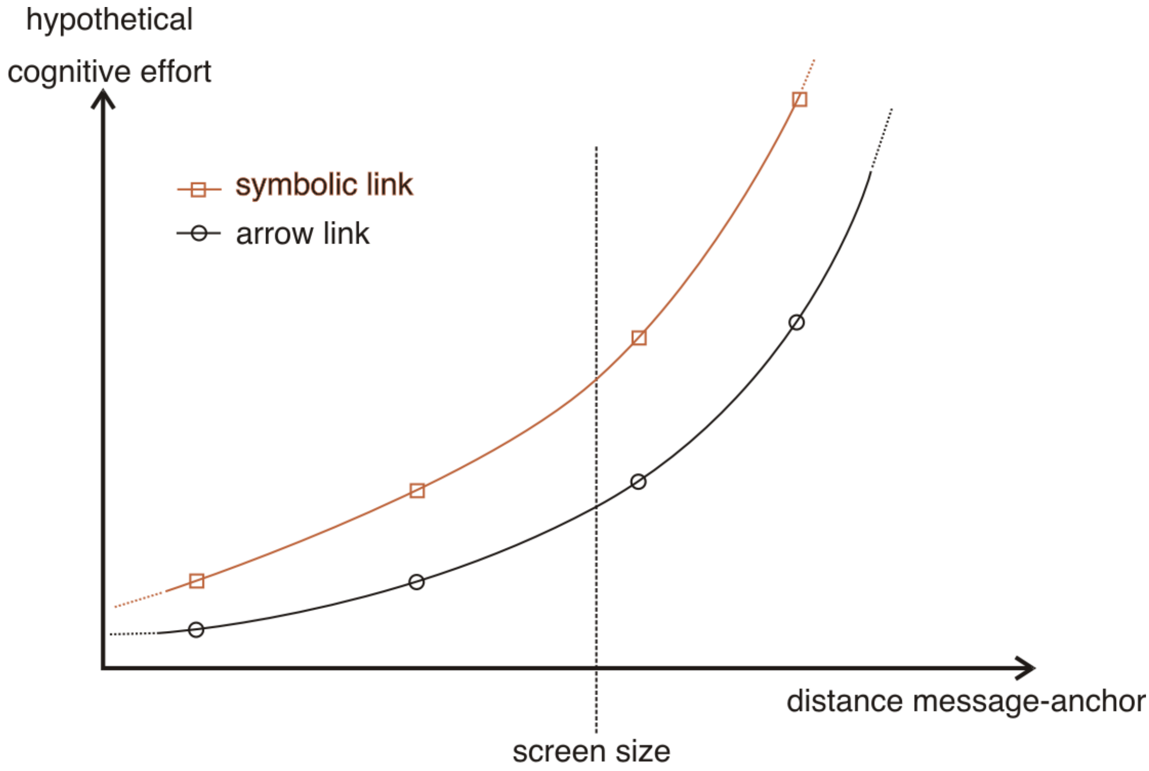


Figure 10-3: Relation between the distance of the message from its anchor point and the contive effort necessary to understand the message. When the distance between the message and its anchor goes beyond the size of the screen, then the user is obliged to scroll the map with an extra cognitive effort required

`M` and `X` must be within a certain visual distance beyond which the user can incur the *split-attention effect* studied by Chandler and Sweller (1992). These authors demonstrated the the shortest visual distance between diagrams and the relative explicating text the better learning performance that students had using these materials. Following this principle, we could have expected the best performance from ShoutSpace, where the distance between the message and the anchor was minimal. However, this was not the case. This thesis suggests that reference does not require necessarily proximity as visual arrows might reduce well the effort of going from the text to the map without incurring in visual cluttering of the map.

Moreover, going from the message to the anchor point should not require scrolling the map because this would require more cognitive effort (e.g., the message should be kept in mind while exploring the map to find the anchor point). This relation is captured by the graph of figure 10-3. There, I represent a hypothetical relationship between the distance message-anchor point and the cognitive effort that is required to make sense of the message. This slope of this curve becomes more steep when the distance increases beyond screen size. The user will be forced to scroll the map thereby losing any visual alignment between the text and their reference points on the map. Finally the cognitive effort required to understand arrow-links may be generally smaller than symbolic links, and this difference less significative at shorter distances. Although this thesis provides clues that reveal this relation, the trade-off between proximity and split attention for referencing should be proved by further research.

10.4.3 Gaze and deictic gestures are intertwined communication mechanisms

Indicating and looking are two intertwined mechanisms that affect collaborative work. When co-located, deictic gestures are used to disambiguate referential expressions, and gazes have the dual nature of perception and communication devices.

At distance, these mechanisms are not possible. Different technological solutions can be implemented to give collaborators the possibility of pointing to portions of the shared display. However, the results reported in this chapter shows that communication mechanisms interact with each other in complex ways. Sustaining only deictic gestures at distance without returning a visual acknowledgment of these acts might not be as efficient as sharing a visual representation of the gaze of the collaboration partners. On the other hand, many scholars have researched the potential benefits of implementing full gaze awareness in video-mediated conversation (Vertegaal, 1999; Monk & Gale, 2002). Unfortunately, we still lack solutions to distinguish when gaze is used as a perception device as opposed to when it is used as a communication device. When face-to-face, people are accustomed to distinguishing between these two, but when at distance, these naturally embodied signals are transposed by technological means and their disambiguation might become tedious[4].

Sharing gazes produced by a participant to perceive elements of the scene might have an overwhelming impact on the collaboration process as not all eye-movements might have comunicative

[4] Of course, eye contact in face-to-face interaction provides different sets of information to which can be gleaned from understanding eye movements over a shared visual representation. For instance gaze can be used to marshal turn-taking. Nevertheless, gaze is indeed used with deictic functions while in presence. Therefore I think it is appropriate to discuss implications that the results presented here might have to the application of gaze awareness in video-mediated communication.

intentions. When co-located, it is our choice to look at our conversation partner's eye to infer where her focus of attention is. It is our choice to look at the person we are addressing to make her aware that our utterances are directed at her. When at distance, it becomes the machine's responsibility to operate this distinction.

Similarly, implementing Explicit Referencing without an acknowledgment feedback might be ineffective. When co-located, we can enrich our conversation with deictic gestures to disambiguate conversation. The speaker can also look at her conversational partners to check whether the gestures that she uses are looked at. When at distance and with the solutions that I tested in this thesis, the emitter of a message cannot know directly whether the recipient has correctly perceived a communication act.

Gaze does not equate to attention

Attention is a selective process through which perceived information is filtered for the limited processing capacity of the brain. Phenomena like inattentional blindness[5] or inattentional amnesia demonstrate the selective nature of attention. Inferring the focus of attention from eye movements is a limited approach. A possible solution would be to combine gaze direction with further evidences of attention coming from different modalities, such as the conversation itself (S. Wood et al., 2006).

Gaze recurrence appears to be a promising marker

The measure that I have adapted in this thesis, namely gaze cross-recurrence seems to be positively related with team performance. These findings open the possibility of using this parameter to measure the quality of collaboration and eventually to offer a regulation feedback. However, I want to caution that this parameter is biased by many factors that need to be considered carefully. As I have shown in section 8.4.1, gaze cross-recurrence is affected by the work strategy chosen by the collaborators (e.g., focusing on a small part of the screen *vs.* a larger area). Additionally, this measure is dependent on the complexity of the display and the way information is encoded in it. A poorly designed representation might place high demands on attention and therefore gaze.

Finally, gaze cross-recurrence is influenced by the symmetry of collaboration. Pairs working in parallel on different sub-tasks might have eye movements that cannot be compared. In fact, this marker becomes meaningful only when the participants are dealing with the same aspect of the task (e.g., looking at the same points of the screen at the same time). This naturally occurs under a management of the interaction following a collaborative paradigm but it is not necessarily the case for pairs adopting a cooperative paradigm.

[5] Inattentional blindness is the phenomenon of not being able an object that is actually there and that can result from a lack of frame of internal frame of reference to perceive the unseen object. See `http://en.wikipedia.org/wiki/Inattentional_blindness`, last retrieved March 2008.

Offering gaze feedback

The introduction of artificial feedback can cause variables which are usually dissociated to become decoupled (S. Wood et al., 2006). Collaborators whose gaze movements might be transferred to their collaboration partners might learn to direct, more intentionally, their sight to specific spots of the interface with specific deictic purposes.

Therefore, more research is needed to understand whether intentional gazes impact collaboration in a different manner compared to naturally embodied glances over the shared scene. As this inputing technique will require training, it will be necessary to perform longitudinal observations as we might register different responses from users after extended periods of use.

10.4.4 Combining linguistic models with eye-tracking and Explicit Referencing

This study offers interesting findings for the design of applications for supporting collaborative work at a distance. Kirk et al. (2007) demonstrated how remote gestures have the potential to reduce the amount of Workers' speech in a physical repair task (a 'Worker' follows the instruction of a 'Helper'). Also in their task, gesturing was associated with a reduction in the occurrence of speech overlaps. Consistently, I have found that Explicit Referencing might reduce the distance between the gaze coupling of the collaborators and that this distance might be associated with repair acts of collaborators' speech. Thus, this study confirms the validity of this interaction mechanism to support collaboration at a distance.

However, this work cautions that the availability of a mechanism like Explicit Referencing does not by itself guarantee the absence of misunderstandings. A possible explanation and implication of this work is related to the way Explicit Referencing was implemented here. In fact, in face-to-face situations, gestures are acknowledged and monitored seamlessly, but in ConcertChat, it was not possible for the emitter to know whether an emitted gesture was seen by the recipient. Gestures and gaze interact in complex ways and it is not trivial to dissociate them.

Similar solutions such as the full gaze awareness (the transmission of the position of the collaborators' sight at a distance), as described by Velichkovsky (1995) or by Monk and Gale (2002) might not be necessary. This work shows how combining information from different communication channels can offer subtle/non persistent solutions to improving the collaboration process without the burden of overwhelming the interaction with a continuous tracking of the position of the eyes that might often be irrelevant. The ability to detect misunderstood messages could be used to offer meaningful information to the collaborators about qualities of their interaction.

This research concerned collaborative annotations of maps. Nonetheless, my findings might be transferred to other domains of application such as collaborative remote text editing. Depending on the complexity of the document or the number of participants, pointing to specific parts of a text might become as problematic as routing objects on a map, and with even less spatial cues than a map can offer.

10.4.5 Implications for eye-tracking research and applications

This work offers also interesting findings for eye-tracking research. The clustering method using the contour function (as in figure 9-2, explained in section 9.3.3) appears to be an effective and deterministic way for detecting the zones of a shared workspace where the user focus on for a period of time longer than a second. Its efficiency should be evaluated against other clustering methods like the mean-shift algorithm developed by Santella and DeCarlo (2004).

In the conclusions of his work on 'inferring intent in eye-based interfaces', Salvucci (1999) argued that "*greater potential arises in the integration of eye movements with other input modalities*". This thesis extends this proposition saying that the integration of eye-tracking and linguistic models yields interesting applications as one modality of communication might help to disambiguate the other. In this regard, the proposed algorithm can be further developed to reduce its dependence on the task-dependent lexicon, as explained in section 10.5.3.

Finally, Ou et al. (2005) showed that with clearly distinct areas of the workspace, it is possible to predict the focus of attention between two partners in a remote collaboration. Using a more complex workspace, I have shown that, given a basic knowledge of the content of the communication exchange and the gazes of the emitter, it is possible to predict where the recipients of a message fixate after the reception of the message.

10.5 Future work

The results of this work suggests ideas on new applications and experiments that complement the findings presented here and foster new research in this domain. The qualitative observations presented in chapter 6 particularly suggested that the production of location-based messages is related to personal benefits for the author. As such, services are not readily available to the general population as they lack an initial mass of information that could generate interest in them, and the mechanism of production should, therefore be bound to another form of reward. A different approach to the matter was suggested by the work of Ludford and her collaborators (Ludford et al., 2007). They designed a location-based reminder service whose content could be reused in a location-based annotation platform. While the main incentive for the user was the value of being reminded of 'to do items' while on the move, the produced content could have been used to populate a location-based annotation service with additional benefits for the users.

Furthermore, the quantitative experiments conducted with ConcertChat and eye-tracking screens suggested the idea of a collaboration platform that could enable both Explicit Referencing and “Gaze Referencing”. The latter corresponds to the idea of allowing the user to enable, **whenever needed and for specific time intervals**, the capture and transfer of her eye-movements to the collaborators for communicative purposes. While Monk and Gale (2002) realized a prototype allowing full gaze awareness, I argue that gaze movements should be transferred to remote collaborators only when needed in relation to the task at hand. A specific interface design should allow the user to choose when and how to use deixis and gaze to resolve a specific reference to the shared workspace (see the mockup presented below at section 10.5.2). Additionally, when a sender produces a message containing a geo-reference, the interface might automatically capture the position of the eyes of the recipient over the provided reference and send this information to the emitter, thus allowing an explicit form of acknowledgment and confirmation.

Finally, the computational model presented in chapter 9 might be improved in several ways (see section 10.5.3). In particular, one of these improvements is the inclusion of a more sophisticated linguistic model than that presented in chapter 9. This would enable the assigning of an acceptability ranking to fixations on the shared map in consideration of the relatum(ii) and the specific relation that is used by the emitter of a message to communicate her intentions. Using such as refined model, it might be possible to predict not only a finer grained index of miscomprehension but also to assign a probability index of comprehension to a message (measuring somewhat its understandability and correctness) produced by the emitter (e.g., emitter writing in her message ‘P200’ but actually looking on ‘P100’).

10.5.1 Location-Based Reminders

The mobile operator has access to a detailed connection log for each mobile phone customer such as all the antennas that were available at a particular time and their signal strength. Using this information, it is possible to localize the customer without any extra specific software running on the phone. I propose to build a location-based reminder system (e.g., being reminded to buy milk when I am close to the grocery store) that locates the user using the network information, instead of using location information provided by ad-hoc hardware or software running on the customer’s phone. Using such system, it would be possible to build a reminder service that does not require special equipment.

Scenario

Miguel wakes up in a cold Monday morning. His wife has already left for work. He is late, plus he has to bring the kids to school. His wife asked him to print some digital pictures that she needs for work the next day. He knows of a shop that offers a print service and that is close to his office. So, he decides to create a ‘location-reminder’ for it. He goes to the PC and connects to a new service of a Spanish telecommunication company that allows the setting up these sorts of reminders related to specific locations. The portal shows the map of Barcelona. He types the name of the shop and immediately, the map adjusts itself to the street of the shop and shows a pinpoint on the right building. He clicks on the pinpoint and he writes a small note for himself to remember what was it is about. He does not specify a time for the reminder to be delivered. On the way to school, his phone vibrates. The system informs Miguel with an SMS that he is passing by a shop that, like the one he selected on the portal, offers the same kind of digital print service. The message contains also the complete address of the shop, business hours, and the phone number. He decides to make a stop to get the prints done.

Initial research questions

An initial research question that this work would focuses on is how to best define a delivery zone. Simple models employ a circular geo-fence that triggers the delivery of the message. However, depending on the architecture of the city in which this system is used and the dynamics of movement of the users, more complex shapes and delivery techniques are needed (Ludford et al., 2006).

Furthermore, this study would focus on how to best optimize the delivery of the reminders based on users' pattern in the urban space. For instance, an user set up a reminder for a certain shopping mall, but she passes by another mall, where two weeks ago she defined a similar reminder. Should her phone ring?

10.5.2 GazeChat

Eye-Chat is an attentive platform for collaboration that allows users to interact at a distance by exchanging text messages, and allowing Explicit Referencing (remote deixis) on a shared workspace. The innovative feature of eye-chat is the capability to monitoring the user's eyes through a webcam attached to a computer and to apply some image recognition and machine learning techniques to infer, with a certain approximation, which parts of the workspace she is looking at. This technique is called low-fidelity eye-tracking and is currently under study (see Pedersen and Spivey 2006). A mockup of the interface of GazeChat is presented in figure 10-4 and was adapted from ConcertChat.

The application also builds a model of the ongoing collaboration by extracting relevant linguistic features from the exchanged messages. The model is used to predict possible misunderstandings between the remote participants. If a conflict is predicted, the spots on which the users were looking are momentarily displayed on the screen in order to allow the participants to repair their different point of view, if any.

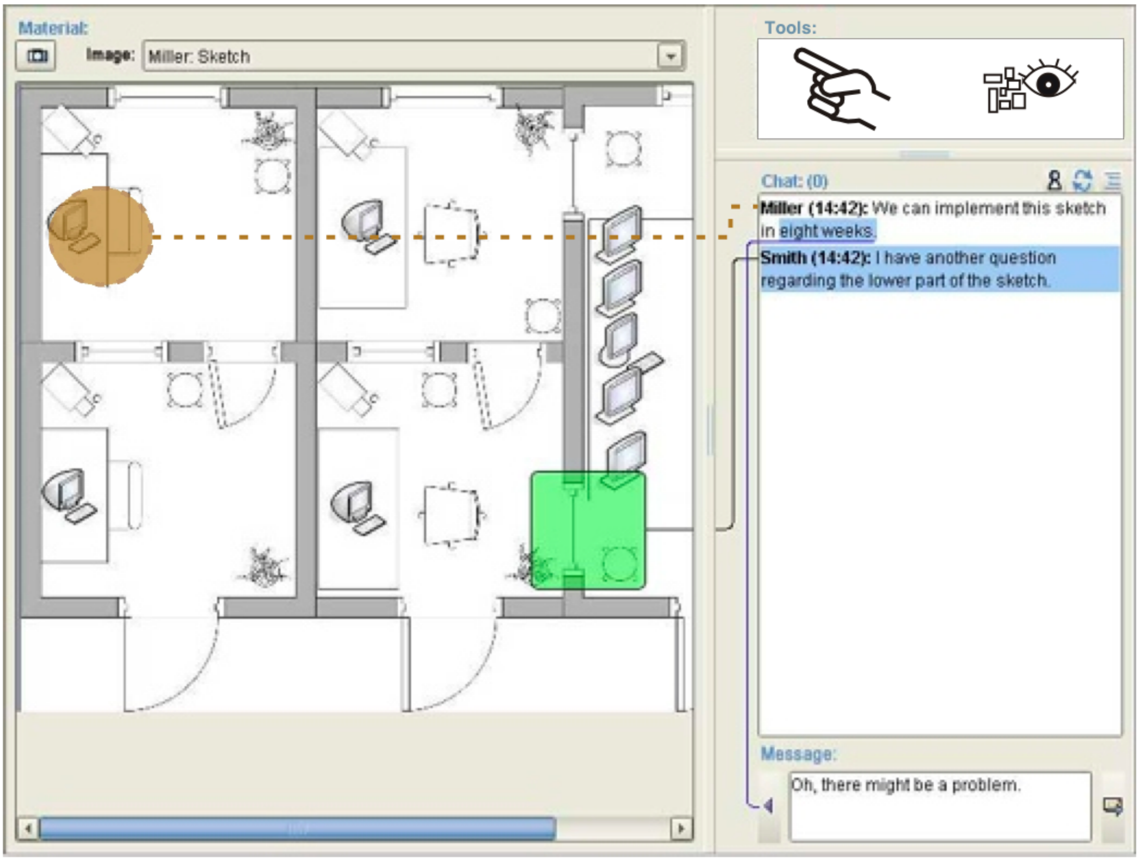


Image: Miller: Sketch

Miller (14:42): We can implement this sketch in eight weeks.

Smith (14:42): I have another question regarding the lower part of the sketch.

Oh, there might be a problem.

Figure 10-4: A mockup of the interface of GazeChat (adapted from ConcertChat). The interface shows two features: the ability of creating Explicit Referencing on the shared workspace (green rectangular selection) and that of enabling for a short time the gaze capture of the local collaborator and the display of this information at the remote sites (brown dashed line with dashed circle)

Scenario

Mike and John are interior designers. Mike is in L.A. while John is working in New York. They have to collaborate to organize an exhibit at MoMA in San Francisco about interactive furniture. They meet online using eye-chat to share ideas. Mike loads the blueprint of the room that they have to furnish for the exhibit. John starts by highlighting with the annotation tool all the electricity plugs available on the walls. Mike chimes in by sketching an initial arrangement of tables and partitioning screens. They comment on each other's work through chat. At one point Mike says: "*we did not consider the natural light coming from the window*", looking at the bottom of the diagram. However, John is looking at the top part of the map when he answers: "*there is enough space between the window and the bench*". At this point, the application overlays to the diagram two semi-transparent circles each of a unique colour assigned to each participant. The two collaborators can easily see that they were looking at two different parts of the screen. Then

Mike explains: *"No, I am not talking about the north wall. I meant the window on the bottom!"*.

Initial research questions

One of the biggest challenges of Eye-Chat is to define the minimal resolution at which eye-tracking might become useful for collaborative work at a distance and under which circumstances (e.g., for which aspects of a task). A possible technique for answering this question would be to employ eye-tracking displays in the experiment and progressively degrading the precision of the tracked locations to measure the resulting effect on collaboration.

A second question would be to understand the interaction between remote deixis and gaze awareness. It would be interesting to test several conditions under which the task can be conducted. Two distinct modalities might be tested: availability of Explicit Referencing (yes/no) and availability of gaze awareness (yes/no). A similar setup was used by Monk and Gale (2002). Finally, gaze awareness might be assisted with an attentive algorithm or in the 'full' condition, continuously displaying eyes position. This will give clear information on the best combination of these tools in supporting collaborative work at a distance and for what kind of task. Additionally, running an experiments with Gaze Chat might help to address some of the limitation of the 'festival' experiment discussed in section 10.2.

10.5.3 A refined linguistic model

To test the algorithm presented in chapter 9, I used an highly detailed semantic map of the EPFL campus (see appendix B, at page 331). Two opposite strategies are possible to ameilorate the algorithm: (a) to simplify the linguisitic model of the algorithm, and (b) to refine this linguistic model.

(a) This map contained a precise definition of all the polygons/shapes that could have been recognized as functional for the task (e.g., buildings, road, crossings, parking lots, etc.). Each polygon was associated with a list of names that were usually used in the chat conversations to refer to it. This information allowed me to test the second hypothesis presented in this paper. Taking the acceptance of this hypothesis, further implementations of this algorithm might use a simpler linguistic model. Simply by knowing how many objects are named in the utterance exchanged by the conversational partners might suffice for the selection of the relevant cluster(s) operated by the clustering module of the procedure.

(b) Conversely, the algorithm can be further improved by taking into account the nature of the relation expressed in the message. If the emitter uses the relation "between" with two *relatii*, then one should expect to find the peak of the cluster in the interspace between the corresponding two polygones. Failure at this test might be attributed to bad encoding of the message, if the emitter is the one not complying with her own writing (fixating elsewhere). Otherwise, this might be

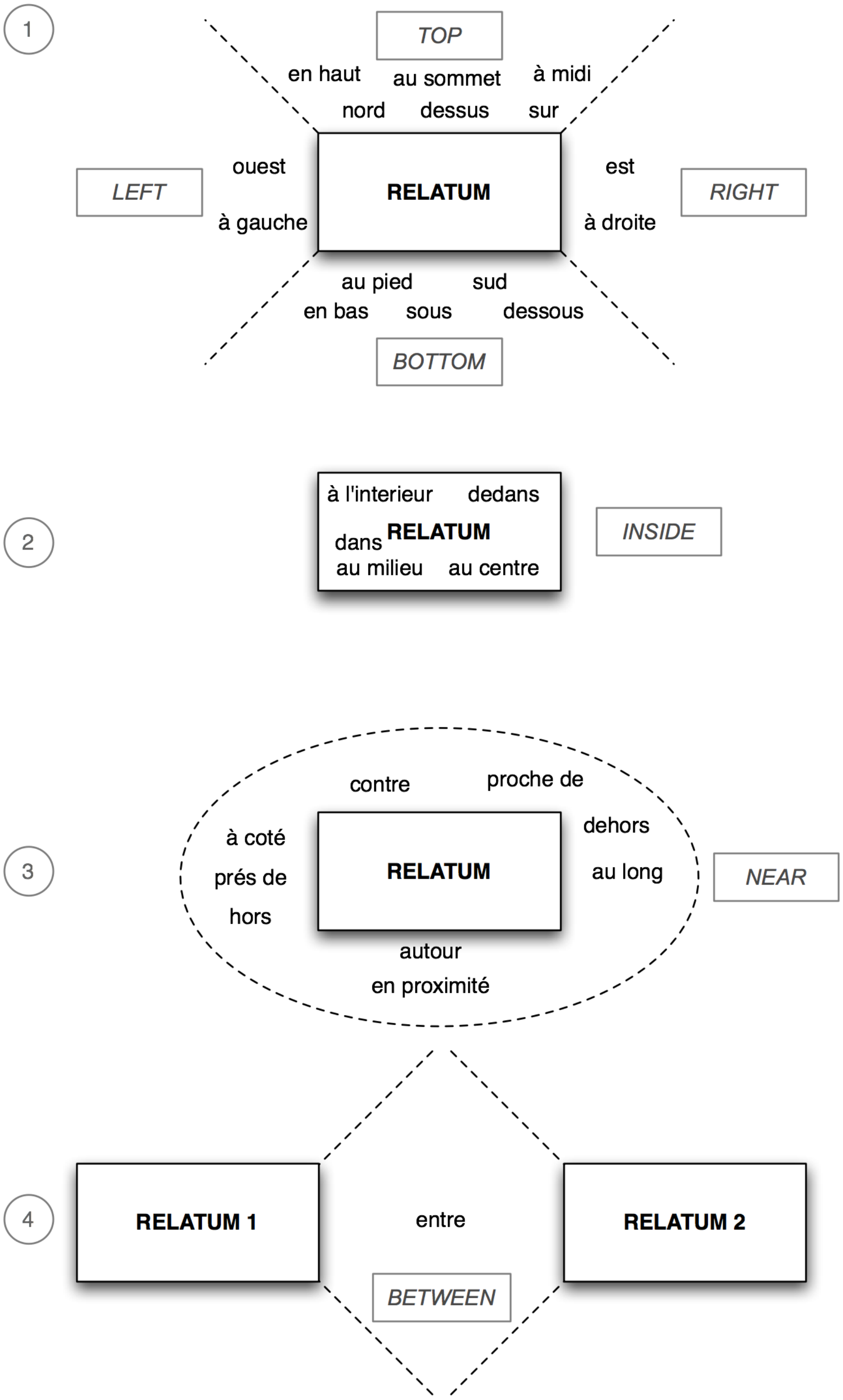


Figure 10-5: Spatial configurations for the seven basic relations extracted from the corpus (labels all capitals in italics). These templates indicate acceptability of fixations in the space where the relatum is located

attributed to a poor understanding of the message, if the recipient is the one fixating further away. Figure 10-5 presents the spatial configurations that might be expected according to which relation is employed by the emitter. I defined a total of seven different spatial relations that corresponds to four different possible tests. The linguistic expressions contained in the schema are the French expressions extracted from the corpus of messages produced in the 'festival' experiment. By using these templates, it might be possible to assign an index of acceptability to the fixations produced by the collaborators during the interaction.

Of course, the proposed spatial configurations are derived from the specific task that was tested. As the orientation of the participants was consistent on the different workstations, they could refer easily with expressions such as "right" or "left", being certain that the partner would have understood correctly. However, these configurations might not be valid for three-dimensional collaborative environments, where participants can modify their perspective and orientation. In these situations, more complex reference frames should be developed (Coventry & Garrod, 2004a).

10.6 Concluding remark

To conclude, this thesis aims at showing that supporting remote collaboration requires finding valid alternatives for communication mechanisms which are effective when collaborators are face-to-face but not available when they are not co-located. I discussed why these alternative mechanisms should not require an increase of bandwidth between the remote sites. The key contribution of this thesis is to show that enabling Explicit Referencing over a shared map is a valid means of achieving effective coordination in collaborative work at a distance for tasks requiring spatial positioning. Furthermore, this work shows that remote deixis is a communication mechanism that is naturally intertwined with gaze and that it is not safe to dissociate these two. Finally, this thesis shows how combining Explicit Referencing, gaze and simple linguistic models might yield interesting results for supporting remote collaboration.

I therefore encourage designers to look for light-weight communication mechanisms that can help remote collaborators to communicate more efficiently without the burden of fracturing their contextual ecologies. In particular, great potential can arise from building attentive interfaces that take into account participants' remote gestures and eye movements.

References

Adams, J., Rogers, B., Hayne, S., Mark, G., Nash, J., & Leifter, L. (2005, January). The effect of a telepointer on student performance and preference. *Computers & Education*, *44*(1), 35-51. Available from `http://dx.doi.org/10.1016/j.compedu.2003.11.002`

Alamargot, D., & Andriessen, J. (2002). Collaborative modeling: Nine recommendations to make a computer supported situation work. In M. Baker, P. Brna, K. Stenning, & A. Tibergien (Eds.), (p. 1-39). Laurence Erlbaum Associated. Available from `http://cogprints.org/3969/`

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representation through speech and gesture. *Psychological Science*, *10*, 327-333.

Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., et al. (1997). Video-mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), (p. 133-155). New Jersey, USA: Lawrence Erlbaum Associates.

André, E., Bosh, G., Herzog, G., & Rist, T. (1987). Artificial intelligence ii: Methodology, systems, applications. In K. Jorrand & L. Sgurev (Eds.), (p. 375-382). Amsterdam, The Nederlands: North-Holland Publishing Company. Available from `http://www.dfki.de/~flint/papers/aimsa86.pdf`

Argyle, M. (1969). *Social interaction*. London, UK: Methuen.

Argyle, M., & Graham, J. (1977, September). The central europe experiment - looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour*, *1*(1), 6-16. Available from `http://www.springerlink.com/content/n28v5818678316w7/`

Arminen, I. (2006). Social functions of location in mobile telephony. *Personal Ubiquitous Computing*, *10*(5), 319-323. Available from `http://dx.doi.org/10.1007/s00779-005-0052-5`

Baker, M., Hansen, T., Joiner, R., & Traum, D. R. (1999). Collaborative learning: Cognitive and computational approaches. In P. Dillenbourg (Ed.), (pp. 31–63). Amsterdam: Pergamon / Elsevier Science. Available from `http://gric.univ-lyon2.fr/gric5/home/mbaker/webpublications/GG-99.PDF`

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*(51), 1173-1182. Available from `http://davidakenny.net/cm/`

mediate.htm

Bauer, M., Kortuem, G., & Segall, Z. (1999, October 18-19). "where are you pointing at?" a study of remote collaboration in a werable videoconference system. In *Proceedings of the3rd international symposium on wearable computers (iswc '99)* (p. 151-158). San Francisco, CA, USA. Available from http://doi.ieeecomputersociety.org/10.1109/ISWC.1999.806696

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*(6), 372-378.

Beigl, M. (2000). Memoclip: A location-based remembrance appliance. *Personal Ubiquitous Computing*, *4*(4), 230-233. Available from http://dx.doi.org/10.1007/s007790070009

Bekker, M. M., Olson, J. S., & Olson, G. M. (1995). Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *Dis '95: Proceedings of the 1st conference on designing interactive systems* (p. 157-166). Ann Arbor, Michigan, United States: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/225434.225452

Benford, S., Bullock, A. N., Cook, N. L., Harvey, P., Ingram, R. J., & Lee, O. (1993). From rooms to cyberspace: Models of interaction in large virtual computer spaces. *Interacting with Computers*, *1*(2), 217-237.

Bereiter, C. (2002). *Education and mind in the knowledge age*. London, UK: Laurence Erlbaum Associated. Available from http://www.cocon.com/observetory/carlbereiter/

Best, D. J., & Roberts, D. E. (1975). Algorithm as 89: The upper tail probabilities of spearman's rho. *Applied Statistics*, *24*, 377-379.

Blades, M., Blaut, J. M., Darvizeh, Z., Elguea, S., Sowden, S., Soni, D., et al. (1998). A cross-cultural study of young children's mapping abilities [doi:10.1111/j.0020-2754.1998.00269.x]. *Transactions of the Institute of British Geographers*, *23*(2), 269-277. Available from http://www.blackwell-synergy.com/doi/abs/10.1111/j.0020-2754.1998.00269.x

Bly, S. A. (1988). A use of drawing surfaces in different collaborative settings. In *Proceedings of the conference on computer-supported cooperative work* (p. 250-256). ACM Press, New York, NY, USA.

Bly, S. A., Harrison, S. R., & Irwin, S. (1993). Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM*, *36*(1), 28-46. Available from http://doi.acm.org/10.1145/151233.151235

Bly, S. A., & Minneman, S. L. (1990). Commune: a shared drawing surface. In *Proceedings of the acm sigois and ieee cs tc-oa conference on office information systems* (p. 184-192). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/91474.91514

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding*. Unpublished doctoral dissertation, Department of Psychology, Stanford University, Stanford, CA, USA. Available from http://wwwlib.umi.com/dxweb/details?doc_no=2583380

Brennan, S. E. (2004). World situated language use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action traditions. In J. Trueswell & M. Tanenhaus (Eds.), (p. 95-130). Cambridge, MA, USA: MIT Press. Available from http://www.psychology.sunysb.edu/sbrennan-/papers/brennan2004.pdf

Brewer, M. (2000). Handbook of research methods in social and personality psychology. In H. Reis & C. Judd (Eds.), (p. 3-16). Cambridge, UK: Cambridge University Press.

Brown-Schmid, S., Campana, E., & Tanenhaus, M. K. (2005). World-situated language processing: Bridging the language as product and language as action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), (chap. Real-time reference resolution by naïve participants during a task-based unscripted conversation). MIT Press. Available from http://www.bcs.rochester.edu/people/ecampana/Papers/2002_CogSciBrownSchmidt.pdf

Brugman, C., & Lakoff, G. (1988). Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology and artificial intelligence. In G. W. Cottrell, S. Small, & M. K. Tanenhaus (Eds.), (p. 477-508). San Matteo, CA, USA: Morgan Kaufmann.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. Oxford, UK: W. W. Norton.

Buckingham-Shum, S., & Sumner, T. (2001, February). Jime: An interactive journal for interactive media. *First Monday*, *6*(2). Available from http://www.firstmonday.org/issues/issue6_2/buckingham_shum/index.html

Burrell, J., & Gay, G. K. (2002). E-graffiti: evaluating real-word use of a context-aware system. *Interacting with Computers*(14), 301-312. Available from http://www.hci.cornell.edu/LabArticles/Egraffiti_Burrell.pdf

Buxton, W. A. S., & Moran, T. P. (1990). Multi-user interfaces and applications. In S. Gibbs & A. A. Verrijn-Stuart (Eds.), (p. 11-34). Amsterdam, The Nederlands: Elsevier.

Campana, E., Beldridge, J., Dowding, J., Hockey, B. A., Remington, R. W., & Stone, L. S. (2001, November 15-16). Using eye movements to determine referents in a spoken dialogue system. In *Electronic proceedings of the workshop on perceptive user interfaces (pui'01).* Orlando, FL, USA. Available from http://www.bcs.rochester.edu/people/ecampana/Papers/2001_PUIC.pdf

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*, 233-246.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1972, December). Studies in interactive communication. i - the effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors*, *14*, 487-509.

Cherny, L. (1999). *Conversation and community: Chat in a virtual world*. Stanford, CA, USA: CSLI

Publications. Available from http://csli-publications.stanford.edu/site/1575861542.html

Cherubini, M., Pol, J. van der, & Dillenbourg, P. (2005, July 5-8). Grounding is not shared understanding: Distinguishing grounding at an utterance and knowledge level. In *Context'05, the fifth international and interdisciplinary conference on modeling and using context.* Paris, France. Available from http://www.i-cherubini.it/mauro/publications/Cherubini_vanderPol_CONTEXT05_dc.pdf

Cherubini, M., Venolia, G., deLine, R., & Ko, A. J. (2007, April 28, May 3). Let's go to the whiteboard: How and why software developers use drawings. In *Proceedings of the sigchi conference on human factors in computing systems (chi2007)* (p. 557 - 566). San Jose, CA, USA: ACM Press. Available from http://doi.acm.org/10.1145/1240624.1240714

Churchill, E. F., Trevor, J., Bly, S. A., Nelson, L., & Cubranic, D. (2000). Anchored conversations: chatting in the context of a document. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 454 - 461). The Hague, The Netherlands. Available from http://portal.acm.org/citation.cfm?id=332475&dl=GUIDE&coll=GUIDE

Clark, H. H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

Clark, H. H. (2003). Pointing: Where language, culture, and cognition meet. In S. Kita (Ed.), (p. 243-268). Mahwah, NJ, USA: Lawrence Erlbaum Associates. Available from http://www-psych.stanford.edu/~herb/2000s/Clark.Pointing.placing.03.pdf

Clark, H. H., & Brennan, S. E. (1991). In l. resnick, j. levine and s. teasley, editors, perspectives on socially shared cognition. In (p. 127-149). Washington: American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*(50), 62-81. Available from http://www-psych.stanford.edu/~herb/2000s/Clark.Krych.04.pdf

Clark, H. H., & Marshall, C. R. (1978). Theoretical issues in natural language processing -2. In D. L. Waltz (Ed.), (p. 57-63). New York, NY, USA: ACM Press.

Clark, H. H., & Marshall, C. R. (1981). Elements of discourse understanding. In A. K. Joshi, I. A. Sag, & B. L. Webber (Eds.), (p. 10-63). Cambridge, UK: Cambridge University Press. Available from http://www-psych.stanford.edu/~herb/

Clark, H. H., & Murphy, G. L. (1982). Language and comprehension. In J.-F. L. Ny & W. Kintsch (Eds.), (p. 287-299). Amsterdam, The Nederlands: North-Holland Publishing Company.

Clark, H. H., & Shaefer, E. F. (1989). Contributing to a discourse. *Cognitive Science*(13), 259-294.

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

Cohn, A. G., Bennett, B., Gooday, J., & Gotts, N. M. (1997). Qualitative spatial representation and reasoning with the region connection calculus. *Geoinformatica*, *1*(3), 275-316. Available from http://citeseer.ist.psu.edu/article/cohn97qualitative.html

Colbert, M. (2007, July). Handbook of mobile human-computer interaction. In J. Lumsden (Ed.), (p. 1-23). London, UK: Idea Group Publishing.

Colston, H. L., & Schiano, D. J. (1995). Looking and lingering as conversational cues in video-mediated communication. In *Chi '95: Conference companion on human factors in computing systems* (p. 278-279). New York, NY, USA: ACM Press. Available from `http://doi.acm.org/10.1145/223355.223673`

Condon, S. L., & Čech, C. G. (2001, January 3-6). Profiling turns in interaction: discourse structure and function. In *Proceedings of the 34th hawai's international conference on system sciences.* Maui, HI, USA: IEEE Computer Society Press. Available from `http://csdl2.computer.org/comp/proceedings/hicss/2001/0981/04/09814034.pdf`

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA, USA: Houghton Mifflin.

Cooper, G. (2002). Wireless world: social and interactional aspects of the mobile age. In B. Brown, R. Green, & R. Harper (Eds.), (p. 19-31). New York, NY, USA: Springer-Verlag New York, Inc.

Coventry, K. R., & Garrod, S. C. (2004a). *Saying, seeing and acting: the psychological semantics of spatial prepositions*. East Sussex, Great Britain: Psychology Press.

Coventry, K. R., & Garrod, S. C. (2004b). Saying, seeing and acting: the psychological semantics of spatial prepositions. In (p. 3-13). East Sussex, Great Britain: Psychology Press.

Coventry, K. R., & Garrod, S. C. (2004c). Saying, seeing and acting: the psychological semantics of spatial prepositions. In (p. 37-70). East Sussex, Great Britain: Psychology Press.

Crabtree, A. (2004). Design in the absence of practice: breaching experiments. In *Dis '04: Proceedings of the 2004 conference on designing interactive systems* (p. 59-68). New York, NY, USA: ACM Press. Available from `http://doi.acm.org/10.1145/1013115.1013125`

Daly-Jones, O., Monk, A. F., & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *Int. J. Hum.-Comput. Stud.*, *49*(1), 21-58.

Debord, G. E. (1955, September). Introduction à une critique de la géographie urbain. *Les Lèves nues*(6).

Dey, A. K., & Abowd, G. D. (2000). Cybreminder: A context-aware system for supporting reminders. In *Huc '00: Proceedings of the 2nd international symposium on handheld and ubiquitous computing* (p. 172-186). London, UK: Springer-Verlag. Available from `http://www.cc.gatech.edu/fce/contexttoolkit/pubs/HUC2000.pdf`

DiEugenio, B., Jordan, P. W., Thomason, R. H., & Moore, J. D. (2000). The agreement process: an empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human Computer Studies*, 1–57. Available from `http://www.cs.uic`

.edu/~bdieugen/

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (p. 1-19). Oxford: Elsevier.

Dillenbourg, P. (2003). Designing biases that augment socio-cognitive interactions. In R. Bromme, R. Hess, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication - and how they may be overcome* (Vol. 5, p. 243-264). Springer Netherlands.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier. Available from http://sir.univ-lyon2.fr/GRIC/GRIC5/Home/mbaker/webpublications/DilBakOmaBla.PDF

Dillenbourg, P., Jermann, P., Schneider, D., Traum, D. R., & Buiu, C. (1997, August 19-22). The design of moo agents: Implications from an empirical cscw study. In *Proceedings of the 8th world conference on artificial intelligence in education (aied'97).* Kobe, Japan. Available from http://tecfa.unige.ch/tecfa/publicat/dil-papers-2/Dil.7.3.21.pdf

Dillenbourg, P., & Traum, D. R. (2006). Sharing solutions: Persistance and grounding in multimodal collaborative problem solving. *The Journal of The Learning Sciences*, *15*(1), 121-151. Available from http://www.leaonline.com/doi/pdf/10.1207/s15327809jls1501_9

Dillenbourg, P., Traum, D. R., & Schneider, D. (1996, September). Grounding in multi-modal task-oriented collaboration. In *Proceedings of the european conference on ai in education* (pp. 415–425). Lisbon, Portugal. Available from http://tecfa.unige.ch/tecfa/research/cscps/euroaied/murder.ps

Dix, A. (1995). Cooperation without (reliable) communication: Interfaces for mobile applications. *Distributed System Engineering*, *3*(2), 171-181. Available from http://www.comp.lancs.ac.uk/~dixa/papers/DSE95/DSE95-mobile.pdf

Donath, J., Karahalios, K., & Viégas, F. B. (1999, January 5-8). Visualizing conversation. In *Proceedings of hicss-32.* Maui, HI: College of Business. Available from http://smg.media.mit.edu/people/Judith/

Dourish, P. (2006, November 4-8). Re-space-ing place: "place" and "space" ten years on. In *Proceedings of the computer supported cooperative work (cscw06)* (p. 299-308). Banff, Alberta, Canada. Available from http://www.ics.uci.edu/%7Ejpd/publications/2006/cscw2006-space.pdf

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the cscw'92 conference on computer-supported cooperative work* (p. 107-114). ACM Press, New York, NY, USA.

Draper, S. W., & Anderson, A. (1991). The significance of dialogue in learning and observing learning. *Computers & Education*, *17*(1), 93-107.

Dyck, J., & Gutwin, C. (2002, April 20-25). Groupspace: a 3d workspace supporting user awareness. In *Proceedings of acm conference on human factors in computing system (chi 2002), extended abstracts* (p. 502-503). Minneapolis, MN, USA.

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamic systems. *Europhysics Letters*(5), 973-977.

Ekbia, H. R., & Maguitman, A. G. (2001). Modeling and using context (proceedings of the conference context 2001, dundee, uk, july). In V. Akman, P. Bouquet, R. Thomason, & R. A. Young (Eds.), (Vol. 2116, p. 156-169). Berlin: Springer.

Finn, K. E. (1997). Video-mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), (p. 3-21). New Jersey, USA: Lawrence Erlbaum Associates.

Frank, A. U., Bittner, S., & Raubal, M. (2001). Spatial information theory - foundations of geographic information science (int. conference cosit 2001, morro bay, usa, september 2001). In D. R. Montello (Ed.), (Vol. 2205, p. 124-139). Derlin, Heidelberg: Springer-Verlag. Available from `ftp://ftp.geoinfo.tuwien.ac.at/frank/af-cosit01-CognizingAgents.pdf`

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*(119), 63-76.

Frohlich, D. M. (1997). Handbook of human-computer interaction. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), (Second, completely revised edition ed., p. 463-488). Amsterdam, The Nederlands: Elsevier Science.

Fuks, H., Pimentel, M. G., & Lucena, C. J. P. de. (2006). R-u-typing-2-me? evolving a chat tool to increase understanding in learning activities. *International Journal of Computer Supported Collaborative Learning*(1), 117-142.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. In *Proceeding of cscw 2000* (p. 21-30). ACM Press, New York, NY, USA. Available from `http://www.cs.cmu.edu/~visual_copresence/BikeStudyCscw2000v18.pdf`

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. I. (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, *19*(3), 273-309. Available from `http://www.leaonline.com/doi/abs/10.1207/s15327051hci1903_3`

Gaver, W., Sellen, A. J., Heath, C., & Luff, P. (1993, April 24-29). One is not enough: Multiple views in a media space. In *Proceedings of interchi'93* (p. 335-341). Amsterdam, The Nederlands: ACM Press. Available from `http://www.goldsmiths.ac.uk/interaction/pdfs/08gaver-etal.MTV.chi93.pdf`

Gergle, D. (2006). *The value of shared visual information for task-oriented collaboration*. Doctoral dissertation - cmu-hcii-06-106, Carnegie Mellon University, Human-Computer Interaction

Institute, Pittsburgh, Pennsylvania, USA. Available from http://www.soc.northwestern.edu/dgergle/pdf/Gergle_Dissertation2006.pdf

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004a, November 6-10). Action as language in a shared visual space. In *Proceedings of the computer supported cooperative work (cscw'04)* (p. 487-496). Chicago, IL, USA. Available from http://www.soc.northwestern.edu/dgergle/pdf/CSCW2004_ActionAsLanguage_Gergle_p487.pdf

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004b). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, *23*(4), 491-517. Available from http://www.cs.cmu.edu/~sfussell/pubs/Manuscripts/JLSP_Gergle%20proofs.pdf

Gergle, D., Kraut, R. E., & Fussell, S. R. (2006, April 22-27). The impact of delayed visual feedback on collaborative performance. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1303-1312). Montréal, Québec, Canada: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/1124772.1124968

Gergle, D., Millen, D. R., Kraut, R. E., & Fussell, S. R. (2004, April 24-29). Persistance matters: Making the most of chat in tightly-coupled work. In *Proceeding of chi2004* (p. 431-438). Vienna, Austria: ACM Press. Available from http://www.cs.cmu.edu/~sfussell/pubs/Manuscripts/CHI04_PersistenceMatters_Gergle.pdf

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA, USA: Houghton Mifflin.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison o high-dimensional and embodied theories of meaning. *Journal of Memory and Language*(43), 379-401. Available from http://dx.doi.org/10.1006/jmla.2000.2714

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think.* Cambridge, MA, USA: Belknap Press.

Golledge, R. G. (Ed.). (1999). *Wayfinding behavior, cognitive mapping and other spatial processes.* London, UK: The Johns Hopkins University Press.

Goodman, J., Brewster, S., & Gray, P. (2004, September). Using field experiments to evaluate mobile guides. In B. Schmidt-Belz & K. Cheverst (Eds.), *Proceedings of hci in mobile guides, workshop at mobile hci.* Glasgow, UK.

Grant, E. R., & Spivey, M. J. (2003, September). Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, *14*(5), 462-466. Available from http://www.cogstud.cornell.edu/spiveylab/guidethought.pdf

Green, G. M. (1987). *Pragmatics and natural language understanding.* New Jersey, USA: Lawrence Erlbaum Associates.

Greenberg, S., Gutwin, C., & Roseman, M. (1996, November 24-27). Semantic telepointers for groupware. In *Proceedings of ozchi'96, sixth australian conference on computer-human* *interaction* (p. 54-61). Hamilton, New Zealand: IEEE Computer Society Press. Available from `http://grouplab.cpsc.ucalgary.ca/papers/1996/96-SemanticTelepointers.OZCHI/cameraready.pdf`

Griffin, Z. M., & Bock, K. (2000, July). What the eyes say about speaking. *Psychological Science*, *11*(4), 274-279.

Grinter, R. E., & Eldridge, M. A. (2001). y do tngrs luv 2 txt msg? In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt, & V. Wulf (Eds.), *Proceedings of the seventh european conference on computer supported cooperative work ecscw'01* (p. 219-238). Bonn, Germany. Available from `http://www.cc.gatech.edu/~beki/c14.pdf`

Grinter, R. E., & Eldridge, M. A. (2003, April 5-10). Wan2tlk?: Everyday text messaging. In *Proceedings of acm conference on human factors in computing system (chi 2003)* (p. 441-448). Fort Lauderdale, Florida, USA. Available from `http://www.cc.gatech.edu/~beki/c24.pdf`

Griswold, W. G., Shanahan, P., Brown, S. W., & Boyer, R. T. (2003). Activecampus: Experiments in community-oriented ubiquitous computing. *IEEE Computer*, *37*(10). Available from `http://www-cse.ucsd.edu/~wgg/Abstracts/ac-handhelds.pdf`

Gutwin, C., & Greenberg, S. (1999). Effects of awareness support on groupware usability. *ACM Transactions on Computer-Human Interaction*, *6*(2), 243-281. Available from `http://hci.usask.ca/publications/1999/effects-tochi.pdf`

Gutwin, C., & Greenberg, S. (2004). The importance of awareness for team cognition in distributed collaboration. In E. Salas, S. Fiore, & J. Cannon-Bowers (Eds.), *Team cognition: Understanding the factors that drives process and performance* (p. 177-201). Washington: APA Press. Available from `http://grouplab.cpsc.ucalgary.ca/papers/2004/04-wa-teamcognition.APABook/wa-teamcog-APABook-2004.pdf`

Guzdial, M. (1997). Information ecology of collaborations in educational settings: Influence of tool. In *Proceedings of the 2nd conference on computer supported collaborative learning (cscl'97)* (p. 83-90). Toronto, CA: University of Toronto.

Hall, E. T. (1966). *The hidden dimension: Man's use of space in public and private*. Garden City, NY, USA: Doubleday.

Hall, G. B., & Leahy, M. G. (2008). Open source approaches to spatial data handling. In G. B. Hall (Ed.), (chap. Design and Implementations of a Map-centered Synchronous Collaboration Tool Using Open Source Components: The MapChat Project). New York, NY, USA: Springer.

Hancock, J. T., & Dunham, P. J. (2001). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes*, *31*(1), 91-110. Available from `http://cucmc.comm.cornell.edu/jth34/publications.php`

Hanna, J. E., & Brennan, S. E. (2007, November). Speaker's eye gaze disambiguate referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596-615.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*(49), 43-61. Available from http://www.psych.upenn.edu/~trueswel/Trueswell_Papers/Hanna_Tanenhaus_Trueswell.pdf

Harnad, S. (1990). The symbol grounding problem. *Physica D*(42), 335-346. Available from http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad90.sgproblem.html

Harrison, S. R., & Dourish, P. (1996). Re-place-ing space: The roles of place and space in collaborative systems. In *Proceedings of the cscw-96.* ACM Press. Available from http://www.ics.uci.edu/~jpd/publications/place-paper.html

Heath, C., & Luff, P. (1991). Disembodied conduct: communication through video in a multimedia office environment. In *Chi '91: Proceedings of the sigchi conference on human factors in computing systems* (p. 99-103). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/108844.108859

Heath, C., Luff, P., Kuzuoka, H., Yamazaki, K., & Oyama, S. (2001, September 16-20). Creating coherent environments for collaboration. In W. Prinz, M. Jarke, Y. Rogers, K. Scmidt, & V. Wulf (Eds.), *Proceedings of ecscw'2001* (p. 119-138). Bonn, Germany: Kluwer Academic Publishers. Available from http://www.grouplab.esys.tsukuba.ac.jp/papers/pdf/ECSCW2001.pdf

Heath, C., Luff, P., & Sellen, A. J. (1997). Video-mediated communication. In K. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), (p. 323-347). Lawrence Erlbaum Associates. Available from http://research.microsoft.com/~asellen/publications/reconfiguring%20media%20space%2097.pdf

Henderson, K. (1999). *On line and on paper: Visual representations, visual culture, and computer graphics in design engineering*. Cambridge, MA, USA: MIT Press.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., & Greenhalgh, C. (1998). Fragmented interaction: establishing mutual orientation in virtual environments. In *Cscw '98: Proceedings of the 1998 acm conference on computer supported cooperative work* (p. 217-226). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/289444.289496

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA, USA: MIT Press.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1986). User centered system design. In D. A. Norman & S. W. Draper (Eds.), (p. 87-124). Hillsdale, N.J., USA: Lawrence Erlbaum Associates.

Hutchins, E. L., & Klausen, T. (1991). Cognition in communication at work. In Y. Engeström & D. Middleton (Eds.), (p. 15-34). New York, USA: Cambridge University Press. Available from http://hci.ucsd.edu/lab/hci_papers/EH1996-1.pdf

Ishii, H., & Kobayashi, M. (1992, May 3-7). Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the sigchi conference on human factors in* *computing systems* (p. 525-532). Monterey, CA, USA: ACM Press. Available from http://doi.acm.org/10.1145/142750.142977

Ishii, H., Kobayashi, M., & Grudin, J. (1993). Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems, 11*(4), 349-375. Available from http://doi.acm.org/10.1145/159764.159762

Ishii, H., & Miyake, N. (1991). Toward an open shared workspace: computer and video fusion approach of teamworkstation. *Commun. ACM, 34*(12), 37-50. Available from http://doi.acm.org/10.1145/125319.125321

Ito, M., & Okabe, D. (2005). Personal, portable intimate: Mobile phones in japanese life. In M. Ito, M. Matsuda, & D. Okabe (Eds.), (p. 1-15). Cambridge, MA, USA: MIT Press. Available from http://www.itofisher.com/mito/mobileemail.pdf

Järvelä, S., & Häkkinen, P. (1999, March 29-31). Web based cases in teaching and learning: Reciprocal understanding and perspective taking in conversation. In *Proceedings of computer assisted learning (cal99).* London, UK.

Jeffrey, P., & Mark, G. (1998). *Constructing social spaces in virtual environments: A study of navigation and interaction* (SICS Technical Report No. T98:02). Stockholm, Sweden: Swedish Institute of Computer Science (SICS). Available from http://www.ece.ubc.ca/~phillipj/papers/ConstructingSocialSpaces.pdf

J. Lévy, M. S. (2005). *Our inhabited space* (Project description poster No. NRP 54). EPFL, CH-1015 Lausanne, Switzerland: Ecole Polytechnique Fédérale de Lausanne, Choros laboratory.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason*. Chicago, IL, USA: University of Chicago Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA, USA: Harvard University Press.

Johnson-Laird, P. N. (1987, March). The mental representation of the meaning of words. *Cognition, 25*(1-2), 189-211.

Jones, M., Buchanan, G., Harper, R., & Xech, P.-L. (2007, April). Questions not answers: a novel mobile search technique. In *Chi '07: Proceedings of the sigchi conference on human factors in computing systems* (p. 155-158). San Jose, California, USA: ACM. Available from http://doi.acm.org/10.1145/1240624.1240648

Jung, Y., Persson, P., & Blom, J. (2005, April 2-7). Dede: Design and evaluation of a context-enhanced mobile messaging system. In *Proceedings of chi 2005.* Portland, Oregon, USA. Available from http://www.perpersson.net/Publications/p206_jung.pdf

Jungnickel, K. (2004). *Urban tapestries: Sensing the city and other stories* (Proboscis Cultural Snapshots No. 9). London, UK: Proboscis. Available from http://proboscis.org.uk/publications/SNAPSHOTS_sensingthecity.pdf

Karsenty, L. (1999). Cooperative work and shared visual context: An empirical study of comprehension problems and in side-by-side and remote help dialogues. *Human-Computer Interaction*, *3*(14), 283-315. Available from http://www.leaonline.com/doi/abs/10.1207/S15327051HCI1403_2

Kasesniemi, E.-L., & Rautiainen, P. (2002). Mobile culture of children and teenagers in finland. In J. E. Katz & M. Aakhus (Eds.), (p. 170-192). New York, NY, USA: Cambridge University Press.

Kenny, D. A., Kashy, D., & Bolger, N. (1998). Handbook of social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), (p. 233-265). Boston, MA, USA: McGraw-Hill.

Kirk, D. S. (2006). *Turn it this way: Remote gesturing in video-mediated communication*. Unpublished doctoral dissertation, Univesrity of Nottingham, Nottingham, UK. Available from http://www.cs.nott.ac.uk/~dsk/DSK-PhDThesisComplete.pdf

Kirk, D. S., Crabtree, A., & Rodden, T. (2005). Ways of the hands. In *Ecscw'05: Proceedings of the ninth conference on european conference on computer supported cooperative work* (p. 1-21). New York, NY, USA: Springer-Verlag New York, Inc. Available from http://www.mrl.nott.ac.uk/~axc/documents/papers/ECSCW05.pdf

Kirk, D. S., & Fraser, D. S. (2006). Comparing remote gesture technologies for supporting collaborative physical tasks. In *Chi '06: Proceedings of the sigchi conference on human factors in computing systems* (p. 1191-1200). Montréal, Québec, Canada: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/1124772.1124951

Kirk, D. S., Rodden, T., & Fraser, D. S. (2007). Turn it this way: grounding collaborative action with remote gestures. In *Chi '07: Proceedings of the sigchi conference on human factors in computing systems* (p. 1039-1048). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/1240624.1240782

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*(1-2), 31-68.

Kirsh, D., & Maglio, P. (1994). On distinguish between epistemic and from pragmatic action. *Cognitive Science*, *18*(4), 513-549.

Koschmann, T., & LeBaron, C. D. (2003). Reconsidering common ground. In K. Kuutti, E. Karsten, G. Fitzpatrick, P. Durish, & K. Schmidt (Eds.), *Ecscw2003: Proceedings of the eight european conference on computer-supported collaborative work.* Amsterdam: Kluwer Academic Publishing.

Koshman, S. (1996). *Uasbility testing of a prototype visualization-based information retrieval system*. Unpublished doctoral dissertation, University of Pittsburgh.

Kosslyn, S. M., Ball, T. M., & Rieser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*(4), 47-60.

Kottamasu, R. (2007). *Placelogging: Mobile spatial annotation and its potential use to urban planners and designers*. Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA. Available from http://www.rajworks.com/rkthesis.pdf

Krauss, R. M., & Fussell, S. R. (1990). Intellectual teamwork: Social and technological foundations of cooperative work. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), (p. 111-147). Hillsdale, N.J., USA: Lawrence Erlbaum Associates.

Krauss, R. M., & Fussell, S. R. (1991). Perspectives on socially shared cognition. In L. B. Resnick, R. M. Levine, & S. D. Teasley (Eds.), (p. 127-149). Washington DC, USA: APA Press.

Krauss, R. M., & Weinheimer, S. (1966, September). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343-346.

Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Distributed work. In P. Hinds & S. Kiesler (Eds.), (p. 137-162). Cambridge, MA, USA: MIT Press.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, *18*, 13-49. Available from http://www.cs.cmu.edu/~kraut/RKraut.site.files/articles/kraut03-VisualInfoConversationalResource-proof.pdf

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of cscw 2002* (p. 31-40). New York, NY, USA: MIT Press. Available from http://doi.acm.org/10.1145/587078.587084

Kulhavy, R. W., & Stock, W. A. (1996). How cognitive maps are learned and remembered [doi:10.1111/j.1467-8306.1996.tb01748.x]. *Annals of the Association of American Geographers*, *86*(1), 123-145. Available from http://www.blackwell-synergy.com/doi/abs/10.1111/j.1467-8306.1996.tb01748.x

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79-86.

Kuzuoka, H. (1992). Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In *Chi '92: Proceedings of the sigchi conference on human factors in computing systems* (p. 533-540). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/142750.142980

Kuzuoka, H., Ishimoda, G., Mishimura, Y., Suzuki, R., & Kondo, K. (1995, September 11-15). Can the gesturecam be a surrogate? In *Proceedings of the fourth european conference on cscw* (p. 181-196). Stockholm, Sweden. Available from http://www.ecscw.org/1995/12.pdf

Kuzuoka, H., Kosuge, T., & Tanaka, K. (1994). Gesturecam: A video communication system for sumpathetic remote collaboration. In *Cscw '94: Proceedings of the 1994 acm conference on computer supported cooperative work* (p. 35-43). Chapel Hill, North Carolina, United States: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/192844

.192866

Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., & Mitsuishi, M. (2000). Gestureman: a mobile robot that embodies a remote instructor's actions. In *Cscw '00: Proceedings of the 2000 acm conference on computer supported cooperative work* (p. 155-162). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/358916.358986

Kuzuoka, H., & Shoki, H. (1994). Findings from observational studies of spatial workplace collaboration. *Electronics and Communications in Japan*, *77*(8), 58-68.

Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., & Heath, C. (2004). Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability. In *Chi '04: Proceedings of the sigchi conference on human factors in computing systems* (p. 183-190). Vienna, Austria: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/985692.985716

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL, USA: The University of Chicago Press. Available from http://www.press.uchicago.edu

Landau, B., & Jackendoff, R. (1993). "what" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(2), 217-238, 255-265. Available from http://web.jhu.edu/cogsci/people/faculty/Landau/papers/Landau1993WhatWhere.pdf

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, industion and representation of knowledge. *Psychological Review*(104), 211-240. Available from http://lsa.colorado.edu/papers/plato/plato.annote.html

Lane, G., Thelwall, S., Angus, A., Peckett, V., & West, N. (2005). *Urban tapestries: Public authoring, place and mobility* (Project final report). London, UK: Proboscis, UK. Available from http://research.urbantapestries.net

Larkin, J. H., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 65-99.

Laurier, E. (2001). Why people say where they are during mobile phone calls. *Environment and Planning D: Society & Space*, *19*(4), 485-504. Available from http://web.ges.gla.ac.uk/~elaurier/texts/whypeople.pdf

Laurillard, D. (1993). *Rethinking university teaching*. London, UK: Routledge.

Lemay, P. (1999). *The statistical analysis of dynamics and complexity in psychology: a configural approach*. Unpublished doctoral dissertation, Political and Social Sciences, University of Lausanne, Lausanne, Switzerland. Available from http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/THX-Doctorat.html

Lemmelä, S.-M., & Korhonen, H. J. (2007). Finding communication hot spots of location-based postings. In *Chi '07: Chi '07 extended abstracts on human factors in computing systems* (p. 2549-2554). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1240866.1241039

Levelt, W. J. M. (1996). Language and space. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), (p. 77-107). Cambridge, MA, USA: MIT Press. Available from http://hdl.handle.net/2066/15546

Lewis, D. K. (1969). *Convention: a philosophical study*. Cambridge, MA, USA: Harvard University Press.

Liben, L. S. (1999). Development of a mental representation. theories and applications. In I. E. Sigel (Ed.), (p. 297-321). Mahwah, NJ, USA: Erlbaum Press.

Ling, R. (2001). *The social juxtaposition of mobile telephone conversations and public spaces* (Research report No. R&D R 45/2001). Fornebu, Norway: Telenor. Available from http://www.telenor.com/rd/pub/rep01/R_45_2001.pdf

Logan, G. D., & Sadler, D. D. (1996). Language and space. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), (p. 493-530). MIT Press.

Lokuge, I., & Ishizaki, S. (1995, May 7-11). Geospace: An interactive visualization system for exploring complex information spaces. In I. R. Katz, R. L. Mack, L. Marks, M. B. Rosson, & J. Nielsen (Eds.), *Chi95: Human factors in computing systems, chi 95 conference proceedings* (p. 409-414). Denver, Colorado, USA: Addison-Wesley. Available from http://doi.acm.org/10.1145/223904.223959

Ludford, P. J. (2006). Sharing everyday places i go while preserving privacy. In *Chi '06: Chi '06 extended abstracts on human factors in computing systems* (p. 1771-1774). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1125451.1125785

Ludford, P. J., Frankowski, D., Reily, K., Wilms, K., & Terveen, L. (2006). Because i carry my cell phone anyway: functional location-based reminder applications. In *Chi '06: Proceedings of the sigchi conference on human factors in computing systems* (p. 889-898). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1124772.1124903

Ludford, P. J., Priedhorsky, R., Reily, K., & Terveen, L. (2007). Capturing, sharing, and using local place information. In *Chi '07: Proceedings of the sigchi conference on human factors in computing systems* (p. 1235-1244). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1240624.1240811

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: Creating environments for collaboration. *Human-Computer Interaction*, *18*(1&2), 51-84. Available from http://www.leaonline.com/doi/abs/10.1207/S15327051HCI1812_3

Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. In *Proceeding of the cognitive science society*

(p. 603-608). Hillsdale, N.J., USA: Erlbaum Press. Available from http://locutus.ucr.edu/reprintPDFs/lba96csp.pdf

Lynch, K. (1964). *L'immagine della città*. Padova: Marsilio Editori.

M., B. J., & D., S. (1971). Studies of geographic learning 1 [doi:10.1111/j.1467-8306.1971.tb00790.x]. *Annals of the Association of American Geographers*, *61*(2), 387-393. Available from http://www.blackwell-synergy.com/doi/abs/10.1111/j.1467-8306.1971.tb00790.x

Maglio, P. P., Barrett, R., Campbell, C. S., & Selker, T. (2000). Suitor: an attentive information system. In *Iui '00: Proceedings of the 5th international conference on intelligent user interfaces* (pp. 169–176). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/325737.325821

Mark, D. M., Freksa, C., Hirtle, S. T., Lloyd, R., & Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Infromation Science*, *13*(8), 747-774. Available from http://www.ingentaconnect.com/search/expand?pub=infobike://tandf/tgis/1999/00000013/00000008/art00002

Marmasse, N., & Schmandt, C. (2000, July). Location-aware information delivery with commotion. *Lecture Notes in Computer Science*, *1927*, 157-171.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*(82), 715-726.

McCarthy, J., & Monk, A. F. (1994). Measuring the quality of computer-mediated communication. *Behavior & Information Technology*, *5*(13), 311-319. Available from http://www.informaworld.com/index/776490963.pdf

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology*(18), 87-121.

McNeill, D. (1992). *Hand and mind. what gestures reveal about thought*. Chicago, IL, USA: University of Chicago Press.

McNeill, D., & Levy, E. (1982). Speech, place and action: Studies in deixis and related topics. In R. J. Jarvella & W. Klein (Eds.), (p. 46). Wiley.

Meyer, T. (2004, Saturday 26th of June). A chacun son guide de la ville sur son mobile. *Le Temps, Switzerland*, p. 16. Available from www.letemps.ch

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London, UK: SAGE Publications.

Milgram, S. (1976). Environmental psychology: People and their physical settings. In H. M. Proshansky, W. H. Ittelson, & L. G. Rivlin (Eds.), (2nd ed., pp. 104–124). New York, USA: Holt, Rinehart and Winston.

Minneman, S. L., & Bly, S. A. (1991). Managing à trois: A study of a multi-user drawing tool in distributed design work. In *Proceedings of the chi'91 conference on human factors in computing* *systems* (p. 217-224). New Orleans, Louisiana, United States: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/108844.108893

Moed, A. (2002). *Annotate space: Interpretation and storytelling on location*. Unpublished master's thesis, Interactive Telecomminications Program, New York University, New York, NY, USA.

Monk, A. F., & Gale, C. (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, *33*(3), 257-278. Available from http://www.leaonline.com/doi/abs/10.1207/S15326950DP3303_4

Morgan, R. (2005). Neighborhood report. new york up close. post-its for passers-by. *New York Times, The City Weekly Desk, Late Edition - Final*(November 13), Section 14, Page 7, Column 3. Available from http://charmandrigor.com/clips/nyt-postits.html

Mühlpfordt, M., & Wessner, M. (2005, May 30 - June 4). Explicit referencing in chat supports collaborative learning. In *Proceedings of the computer supported collaborative learning 2005* (p. 460-469). Taipei, Taiwan. Available from http://portal.acm.org/citation.cfm?id=1149293.1149353

Nardi, B. A. (2005). Beyond bandwidth: Dimensions of connection in interpersonal communication. *Computer Supported Cooperative Work (CSCW)*, *14*(2), 91-130. Available from http://dx.doi.org/10.1007/s10606-004-8127-9

Newcombe, N., & Liben, L. S. (1982, August). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, *34*(1), 46-58.

Nova, N. (2005). A review of how space affords socio-cognitive processes during collaboration. *Psychonology*, *3*(2), 118-148.

Nova, N. (2007). *The influence of location awareness on computer-supported collaboration*. n. 3769, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. Available from http://biblion.epfl.ch/EPFL/theses/2007/3769/EPFL_TH3769.pdf

Nova, N., Girardin, F., & Dillenbourg, P. (2005, November 28-30). 'location is not enough!': an empirical study of location-awareness in mobile collaboration. In *Proceeding of the third ieee international workshop on wireless and mobile technologies in educations* (p. 21-28). Tokushima, Japan: IEEE Press. Available from http://craftsrv1.epfl.ch/publications/2005/paper_wmte2005.pdf

Nova, N., Wehrle, T., Goslin, J., Bourquin, Y., & Dillenbourg, P. (2003, September). The impacts of awareness tools on mutual modelling in a collaborative video-game. In J. Favela & D. Decouchant (Eds.), *Proceedings of the 9th international workshop on groupware* (p. 99-108). Autrans, France. Available from http://tecfa.unige.ch/perso/staf/nova/paper_criwg.pdf

O'Conaill, B., & Whittaker, S. (1997). Video-mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), (p. 107-131). New Jersey, USA: Lawrence Erlbaum Associates.

Oh, A., Fox, H., Kleek, M. V., Adler, A., Gajos, K., Morency, L.-P., et al. (2002). Evaluating look-to-talk: a gaze-aware interface in a collaborative environment. In *Chi '02: Chi '02 extended abstracts on human factors in computing systems* (p. 650-651). New York, NY, USA: ACM Press. Available from `http://doi.acm.org/10.1145/506443.506528`

Olson, J. S., Olson, G. M., & Meader, D. K. (1995, May 7-11). What mix of video and audio is useful for small groups doing remote real-time design work? In *Proceedings of conference on humans factors and computing systems (chi)* (p. 362-368). Denver, Colorado, USA: ACM Press. Available from `http://sigchi.org/chi95/proceedings/papers/jso_bdy.htm`

Ou, J., Chen, X., Fussell, S. R., & Yang, J. (2003, November 2-8). Dove: Drawing over video environment. In *Proceedings of multimedia'03: Demonstrations.* Berkeley, CA, USA. Available from `http://www.cs.cmu.edu/~gestures/papers/de007-ou_v1.pdf`

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., & Yang, J. (2003). Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In *Icmi '03: Proceedings of the 5th international conference on multimodal interfaces* (p. 242-249). New York, NY, USA: ACM Press. Available from `http://doi.acm.org/10.1145/958432.958477`

Ou, J., Oh, L. M., Fussell, S. R., Blum, T., & Yang, J. (2005, October 4-6). Analyzing and predicting focus of attention in remote collaborative tasks. In *Proceedings of the 7th international conference on multimodal interfaces icmi '05* (p. 116-123). Trento, Italy: ACM Press. Available from `http://doi.acm.org/10.1145/1088463.1088485`

Ou, J., Oh, L. M., Yang, J., & Fussell, S. R. (2005). Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task. In *Chi '05: Proceedings of the sigchi conference on human factors in computing systems* (p. 231-240). New York, NY, USA: ACM. Available from `http://doi.acm.org/10.1145/1054972.1055005`

Paivio, A. (1990). *Mental representations: A dual coding approach.* New York, NY, USA: Oxford University Press.

Pedersen, B., & Spivey, M. J. (2006). Offline tracking of eyes and more with a simple webcam. In *Proceedings of the workshop "what have eye movements told us so far, and what is next?", 28th annual meeting of the cognitive science society, cogsci2006 proceedings.* Available from `http://www.people.cornell.edu/pages/bp64/PedersenSpivey.pdf`

Persson, P., & Fagerberg, P. (2002). *Geonotes: a real-use study of a public location-aware community system* (Technical Report No. T2002:27). Sweden: SICS. Available from `http://www.sics.se/~petra/techReport.pdf`

Phillips, B. (2000). Should we take turns?: a test of cmc turn-taking formats. In *Chi '00 extended abstracts on human factors in computing systems* (p. 341-342). New York, NY, USA: ACM Press. Available from `http://doi.acm.org/10.1145/633292.633497`

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral* *and Brain Sciences*(27), 169-226.

Pickering, M. J., & Garrod, S. (2006, October). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228.

Pimentel, M. G., Fuks, H., & Lucena, C. J. P. de. (2003, June). Co-text loss in textual chat tools. In *4th international and interdisciplinary conference on modeling and using context (context2003)* (Vol. LNAI 2680, p. 483-490). Stanford, CA, USA. Available from `http://www.springerlink.com/content/xt15fjgrptuegutk/`

Pol, J. van der, Admiraal, W., & Simons, P. R. J. (2006a, August). The affordance of anchored discussion for the collaborative processing of academic texts. *Computer Supported Collaborative Learning*(1), 339-357. Available from `http://www.springerlink.com/content/p1905251815r8882/`

Pol, J. van der, Admiraal, W., & Simons, P. R. J. (2006b). Context enhancement for co-intentionality and co-reference in asynchronous cmc. *Journal of artificial intelligence & society*, *3*(20), 301-313. Available from `http://www.uu.nl/content/b56k1gkn54255422.pdf`

Pomplun, M., Ritter, H., & Velichkovsky, B. M. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*(25), 931-948. Available from `http://citeseer.ist.psu.edu/pomplun95disambiguating.html`

Proshansky, H. M., Ittelson, W. H., & Rivlin, L. G. (1970). Environmental psychology: People and their physical settings. In H. M. Proshansky, W. H. Ittelson, & L. G. Rivlin (Eds.), (p. 177-181). New York, NY, USA: Holt, Rinehart and Winston.

Purnell, K. N., Solman, R. T., & Sweller, J. (1991). The effects of technical illustrations on cognitive load. *Instructional Science*(20), 443-462. Available from `http://www.springerlink.com/content/t4n7836218217171/`

Qvarfordt, P., Beymer, D., & Zhai, S. (2005, September 12-16). Realtourist: A study of augmenting human-human and human-computer dialogue with eye-gaze overlay. In M. F. Costabile & F. Paternò (Eds.), *Proceedings of interact 2005* (Vol. 3585, p. 767-780). Rome, Italy: Springer. Available from `http://www.springerlink.com/content/ac7q9de8je0dj4vj`

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*(7), 226-231.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273-298. Available from `http://www.psych.uchicago.edu/~regier/papers/avs-jepg.pdf`

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*(29), 1045-1060. Available from `http://psych.ucsc.edu/eyethink/publications_assets/RichardsonDale2005.pdf`

Robertson, T. (1997). Cooperative work and lived cognition: a taxonomy of embodied actions. In *Ecscw'97: Proceedings of the fifth conference on european conference on computer-supported cooperative work* (p. 205-220). Norwell, MA, USA: Kluwer Academic Publishers. Available from http://research.it.uts.edu.au/idwop/downloads/RobertsonECSCW1997.pdf

Robson, C. (2002). *Real world research* (2nd edition ed.). Oxford, UK: Blackwell.

Rochelle, J., & Teasley, S. D. (1995). Computer supported collaborative learning. In C. O'Malley (Ed.), (p. 69-197). Berlin, Germany: Springer-Verlag.

Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: an attributional bias in self-perception and social perception processes. *Journal of Experimental Social Psychology*(13), 279-301.

Ryan, N. (2005, February). Smart environments for cultural heritage. In *Reading the historical spatial information in the world, 24th international symposium.* Kyoto, Japan. Available from http://www.mobicomp.org/FieldMap

Salgado, M., & Diaz-Kommonen, L. (2006, March 1). Visitors' voices. In J. Trant & D. Bearman (Eds.), *Museums and the web 2006: Proceedings.* Toronto, CA. Available from http://www.archimuse.com/mw2006/papers/salgado/salgado.html

Salvucci, D. D. (1999, May 15-20). Inferring intent in eye-based interfaces: Tracing eye movements with process models. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 254-261). Pittsburgh, Pennsylvania, USA. Available from http://doi.acm.org/10.1145/302979.303055

Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Etra '04: Proceedings of the 2004 symposium on eye tracking research & applications* (p. 27-34). San Antonio, Texas: ACM Press, New York, NY, USA. Available from http://doi.acm.org/10.1145/968363.968368

Schelling, T. (1960). *The strategy of conflict.* Cambridge, MA, USA: Harvard University Press.

Schiffer, S. (1972). *Meaning.* Oxford, UK: Claredon Press.

Schmid, H. (1994, September). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of international conference on new methods in language processing.* Available from http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.ps.gz

Schneiderman, B. (1982). The future of inteactive systems and the emergence of direct manipulation. *Behaviour and Information Technology*(1), 237-256.

Schneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *IEEE Computer*(16), 57-69.

Schwartz, D. L., & Lin, X. D. (2000). Computers, productive agency, and the effort after shared meaning. *Journal of Computing in Higher Education*, *12*(2), 3-33.

Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-*

Computer Interaction, *10*, 401-444. Available from http://www.leaonline.com/doi/pdfplus/10.1207/s15327051hci1004_2

Sellen, A. J., & Harper, R. (1997). Video-mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), (p. 225-243). New Jersey, USA: Lawrence Erlbaum Associates.

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA, USA: Houghton Mifflin.

Shepard, R. N., & Metzler, J. (1971). Mental rotations of three-dimensional objects. *Science*(171), 701-703.

Shepard, R. N., & Podgorny, P. (1978). Handbook of learning and cognitive processes. In W. K. Estes (Ed.), (Vol. 5, p. 189-237). Hillsdale, N.J., USA: Erlbaum Press.

Smith, I., Consolvo, S., LaMarca, A., Hightower, J., Scott, J., Sohn, T., et al. (2005, May, 8-13). Social disclosure of place: From location technology to communication practices. In *Proceedings of pervasive05.* Munich, Germany. Available from http://www.placelab.org/publications/pubs/pervasive-privacy-2005-final.pdf

Smith, M., Cadiz, J. J., & Burkhalter, B. (2000, New York, NY, USA). Conversation trees and threaded chats. In *Cscw '00: Proceedings of the 2000 acm conference on computer supported cooperative work* (pp. 97–105). Philadelphia, Pennsylvania, United States: ACM Press. Available from http://doi.acm.org/10.1145/358916.358980

Smith, N. (1982). *Mutual knowledge*. New York, NY, USA: Academic Press Ltd.

Sobel, M. E. (1982). Sociological methodology. In (p. 290-312). San Francisco, CA, USA: Jossey-Bass.

Sohn, T., Li, K. A., Lee, G., Smith, I., Scott, J., & Griswold, W. G. (2005, September). Place-its: A study of location-based reminders on mobile phones. In *Ubicomp'05: Seventh international conference on ubiquitous computing* (p. 232-250). Tokyo, Japan. Available from http://www.cse.ucsd.edu/users/wgg/Abstracts/tsohn-placeits-ubicomp05-final.pdf

Sperber, D., & Wilson, D. (1986). *Relevance. communication & cognition*. Oxford, UK: Blackwell.

Stahl, G. (2000). A model of collaborative knowledge-building. In *Proceedings of the fourth international conferences of the learning sciences (icls).* MI, USA: Ann Arbor. Available from http://www.cis.drexel.edu/faculty/gerry/cscl/papers/ch14.pdf

Stahl, G., Zemel, A., Sarmiento, J., & Cakir, M. (2006, June 27-July 1). Shared referencing of mathematical objects in online chat. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of icls2006, the 7th international conference of the learning sciences* (Vol. 2, p. 716-722). Indiana University, Bloomington, IN: Lawrence Erlbaum Associates.

Suthers, D., Girardeau, L., & Hundhausen, C. (2003). Designing for change. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), (p. 173-182). Amsterdam, The Nederlands: Kluwer Academic Publishers. Available from http://lilt.ics.hawaii.edu/lilt/papers/2003/

Suthers-et-al-CSCL2003.pdf

Suthers, D., & Xu, J. (2002, May 7-11). Kükäkükä: An online environment for artifact-centered discourse. In *Education track of the eleventh world wide web conference (www 2002)* (pp. 472–480). Honolulu, HI, USA. Available from http://www2002.org/CDROM/alternate/252/

Suwa, M., Gero, J., & Purcell, T. (2000). Unexpected discoveries and s-invention of design requirements: Improving vehicles for a design process. *Design Studies*, *18*(4), 539-567.

Suwa, M., & Tversky, B. (1997). What architects and students perceive in their sketches: A protocol analysis. *Design Studies*(18), 385-403. Available from http://psychology.stanford.edu/~bt/diagrams/papers/suwaarchitectssee.doc.pdf

Talmy, L. (1983). Spatial orientation: Theory, research and application. In H. L. Pick & L. P. Acredolo (Eds.), (p. 225-282). New York, NY, USA: Plenum. Available from http://linguistics.buffalo.edu/people/faculty/talmy/talmyweb/Volume1/chap3.pdf

Tang, A., Boyle, M., & Greenberg, S. (2004). Display and presence disparity in mixed presence groupware. In *Auic '04: Proceedings of the fifth conference on australasian user interface* (p. 73-82). Darlinghurst, Australia, Australia: Australian Computer Society, Inc. Available from http://portal.acm.org/citation.cfm?id=976320#

Tang, J. C. (1989). *Listing, drawing and gesturing in design: A study of the use of shared workspaces by design teams*. Unpublished doctoral dissertation, Department of Mechanical Engineering. Stanford University, Stanford, CA, USA.

Tang, J. C. (1991). Findings from observational studies of collaborative work. *Int. J. Man-Mach. Stud.*, *34*(2), 143-160.

Tang, J. C., & Isaacs, E. A. (1993). Why do users like video? studies of multimedia-supported collaboration. *Computer-Supported Collaborative Work: An International Journal*, *1*(9), 163-196. Available from http://research.sun.com/techrep/1992/smli_tr-92-5.pdf

Tang, J. C., & Minneman, S. L. (1991a). Videodraw: a video interface for collaborative drawing. *ACM Trans. Inf. Syst.*, *9*(2), 170-184. Available from http://doi.acm.org/10.1145/123078.128729

Tang, J. C., & Minneman, S. L. (1991b). Videowhiteboard: video shadows to support remote collaboration. In *Chi '91: Proceedings of the sigchi conference on human factors in computing systems* (p. 315-322). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/108844.108932

Tatar, D. G., Foster, G., & Bobrow, D. G. (1991). Designing for conversation: Lessons from cognoter. *International Journal of Man-Machine Studies*(34), 143-160.

Taylor, H. A., & Tversky, B. (1992a). Descriptions and depictions of environments. *Memory and Cognition*(20), 483-496.

Taylor, H. A., & Tversky, B. (1992b). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*(31), 261-292.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*(3), 371-391. Available from http://www.sciencedirect.com/science/article/B6WK4-45MG3KJ-X/2/7d8d24633b964168cf1af429b54b7ee2

Tester, J. (2004, October 22). *Why people will geo-annotate physical space.* blog post. Retrieved March 2008, from http://future.iftf.org/2004/10/why_people_will.html

Traum, D. R. (1999, November). Computational models of grounding in collaborative systems. In *working notes of aaai fall symposium on psychological models of communication* (pp. 124–131). Available from http://www.ict.usc.edu/~traum/Papers/psych.ps

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd Edition ed.). Cheshire, Connecticut, USA: Graphics Press.

Tungare, M., Burbey, I., & Pérez-Quinones, M. A. (2006). Evaluation of a location-linked notes system. In *Acm-se 44: Proceedings of the 44th annual southeast regional conference* (p. 494-499). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/1185448.1185557

Tversky, B. (1977). Memory from emotion and everyday events. In N. Stein, P. Ornstein, B. Tversky, & C. Brainerd (Eds.), (p. 181-208). Mahwah, NJ, USA: Erlbaum Press.

Tversky, B. (1993). Spatial information theory - a theoretical basis for gis (int. conference cosit 1993, elba, italy). In A. U. Frank & I. Campari (Eds.), (Vol. 716, p. 14-24). Berlin, Germany: Springer. Available from http://www.geo.unizh.ch/cosit03/tversky-COSIT93.pdf

Tversky, B. (1996). Language and space. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), (p. 463-491). Cambridge, MA, USA: MIT Press.

Tversky, B. (1998). Theories of memory ii. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), (p. 259-275). Hove, East Sussex, GB: Psychological Press.

Tversky, B. (2001). Spatial schemas and abstract thought. In M. Gattis (Ed.), (p. 79-111). Cambridge, MA, USA: MIT Press.

Tversky, B. (2005). The cambridge handbook of thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), (p. 209-240). Cambridge, MA, USA: Cambridge University Press. Available from http://www-psych.stanford.edu/~bt/space/papers/visuospatialholyoak04%20.pdf

Tversky, B., Heiser, J., Lozano, S., MacKenzie, R., & Morrison, J. B. (2007). Learning with animation. In R. Lowe & W. Schnotz (Eds.), (p. 263-286). Cambridge, MA, USA: Cambridge University Press.

Tversky, B., & Lee, P. U. (1998). How space structure language. In *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (Vol. Volume 1404/1998, p. 157-175). Heidelberg: Springer-Verlag Heidelberg. Available from http://www-psych

.stanford.edu/~bt/space/papers/spacestructureslanguage.doc.pdf

Tversky, B., Suwa, M., Agrawala, M., J, H., Stolte, C., Hanrahan, P., et al. (2003). Human behavior in design: Individuals, teams, tools. In U. Lindemann (Ed.), (p. 79 - 86). Berlin, Germany: Springer. Available from http://www-psych.stanford.edu/~bt/diagrams/papers/sketches-tversky-suwa-etal.pdf

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA, USA: MIT Press.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology*, *1*(57A), 165-182. Available from http://www.psychology.nottingham.ac.uk/staff/Geoff.Underwood/Underwood_Papers/QJEP.2004.pdf

Uttal, D. H. (2000, August). Seeing the big picture: Map use and the development of spatial cognition. *Developmental Science*, *3*(3), 247–264. Available from https://www.depot.northwestern.edu/projects/wcas/psych/uttallab/WebPublications/uttal2000.pdf

Vandeloise, C. (1991). *Spatial prepositions: a case study from french*. The University of Chicago Press. Available from http://www.amazon.com/Spatial-Prepositions-Case-Study-French/dp/0226847284

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, *3*(2), 199-222.

Vertegaal, R. (1999). The gaze groupware system: mediating joint attention in multiparty communication and collaboration. In *Chi '99: Proceedings of the sigchi conference on human factors in computing systems* (p. 294-301). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/302979.303065

Vertegaal, R., Slagter, R., Veer, G. van der, & Nijholt, A. (2001, March 31-April 4). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 301-308). Seattle, WA, USA. Available from http://doi.acm.org/10.1145/365024.365119

Viégas, F. B., Wattenberg, M., & Dave, K. (2004, April 24-29). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of acm conference on human factors in computing system (chi 2004)* (p. 575-582). Vienna, Austria. Available from http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf

Wang, J., & Canny, J. (2006). End-user place annotation on mobile devices: a comparative study. In *Chi '06: Chi '06 extended abstracts on human factors in computing systems* (p. 1493-1498). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1125451.1125725

Weiss, G., & Dillenbourg, P. (1999). Collaborative-learning: Cognitive and computational approaches. In P. Dillenbourg (Ed.), (p. 64-80). Oxford, UK: Elsevier.

Wellner, P. (1993). Interacting with paper on the digitaldesk. *Commun. ACM*, *36*(7), 87-96. Available from http://doi.acm.org/10.1145/159544.159630

Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications: Theory and design implications. *International Journal of Human-Computer Studies*(42), 501-529.

Whittaker, S., Brennan, S. E., & Clark, H. H. (1991). Co-ordinating activity: An analysis of interaction in computer-supported co-operative work. In *Proceedings of the chi'91 conference on human factors in computing systems* (p. 361-367). ACM Press, New York, NY, USA.

Whittaker, S., Geelhhoed, E., & Robinson, E. (1993, November). Shared workspaces: how do they work and when are they useful? *International Journal of Man-Machine Studies*, *39*(5), 813-842. Available from http://dx.doi.org/10.1006/imms.1993.1085

Wood, D. (1992). *The power of maps*. New York, NY, USA: Guilford Press.

Wood, S., Cox, R., & Cheng, P. (2006, February). Attention design: Eight issues to consider. *Computers in Human Behavior*, *22*, 588-602. Available from http://eidetic.ai.ru.nl/egon/education/IntroCE/literature/attention_emotion/Wood06-Attention_design.pdf

Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., Kato, H., Suzuki, H., et al. (1999, September 12-16). Gesturelaser and gesturelaser car: Development of an embodied space to support remote instruction. In S. Bodker, M. Kyng, & K. Schmidt (Eds.), *Proceedings of the sixth european conference on computer-supported cooperative work (ecscw'99)* (p. 239-258). Copenhagen, Denmark: Kluwer Academic Publishers.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (magic) pointing. In *Chi '99: Proceedings of the sigchi conference on human factors in computing systems* (p. 246-253). New York, NY, USA: ACM Press. Available from http://doi.acm.org/10.1145/302979.303053

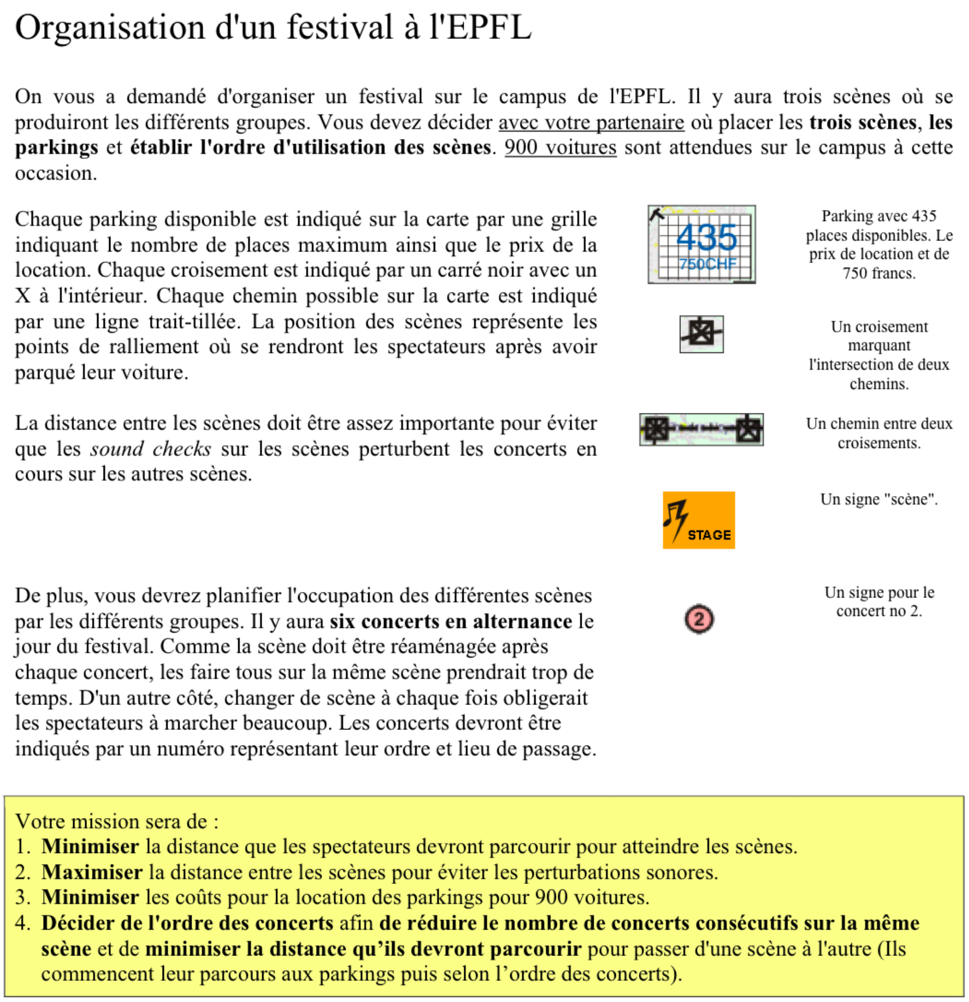
Zhou, C., Ludford, P. J., Frankowsky, D., & Terveen, L. (2005, April 2-7). An experiment in discovering personally meaningful places from location data. In A. C. Machinery (Ed.), *Proceedings chi2005* (p. 2029-2032). Portland, Oregon, USA. Available from http://portal.acm.org/ft_gateway.cfm?id=1057084&type=pdf&coll=GUIDE&dl=GUIDE&CFID=45874047&CFTOKEN=55447131

Index of Authors

Index of Keywords

Appendix A

Materials used for the experiments



Organisation d'un festival à l'EPFL

On vous a demandé d'organiser un festival sur le campus de l'EPFL. Il y aura trois scènes où se produiront les différents groupes. Vous devez décider avec votre partenaire où placer les **trois scènes**, **les parkings** et **établir l'ordre d'utilisation des scènes**. 900 voitures sont attendues sur le campus à cette occasion.

Chaque parking disponible est indiqué sur la carte par une grille indiquant le nombre de places maximum ainsi que le prix de la location. Chaque croisement est indiqué par un carré noir avec un X à l'intérieur. Chaque chemin possible sur la carte est indiqué par une ligne trait-tillée. La position des scènes représente les points de ralliement où se rendront les spectateurs après avoir parqué leur voiture.

La distance entre les scènes doit être assez importante pour éviter que les *sound checks* sur les scènes perturbent les concerts en cours sur les autres scènes.

De plus, vous devrez planifier l'occupation des différentes scènes par les différents groupes. Il y aura **six concerts en alternance** le jour du festival. Comme la scène doit être réaménagée après chaque concert, les faire tous sur la même scène prendrait trop de temps. D'un autre côté, changer de scène à chaque fois obligerait les spectateurs à marcher beaucoup. Les concerts devront être indiqués par un numéro représentant leur ordre et lieu de passage.

435 750CHF	Parking avec 435 places disponibles. Le prix de location et de 750 francs.
☒	Un croisement marquant l'intersection de deux chemins.
☒- - -☒	Un chemin entre deux croisements.
STAGE	Un signe "scène".
②	Un signe pour le concert no 2.

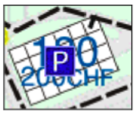
Votre mission sera de :
1. **Minimiser** la distance que les spectateurs devront parcourir pour atteindre les scènes.
2. **Maximiser** la distance entre les scènes pour éviter les perturbations sonores.
3. **Minimiser** les coûts pour la location des parkings pour 900 voitures.
4. **Décider de l'ordre des concerts** afin **de réduire le nombre de concerts consécutifs sur la même scène** et de **minimiser la distance qu'ils devront parcourir** pour passer d'une scène à l'autre (Ils commencent leur parcours aux parkings puis selon l'ordre des concerts).

Figure A-1: Participant instruction sheet used for the 'festival' experiment, page 1

RULES:

(1)

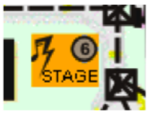
Vous avez à votre disposition 13 signes Parking. Pour sélectionner un parking, il faut poser un signe sur la grille représentant le parking choisi. **Vous devez choisir les parkings de sorte à dépenser le moins d'argent possible en location.**



Ce parking a été choisi.

(2)

Vous devrez également placer les trois scènes sur le plan à l'aide d'un signe. Vous pouvez positionner les scènes sur des emplacements libre (espaces verts-pâle), éventuellement à un croisement, mais PAS sur un bâtiment ni sur un parking. Choisissez soigneusement où les spectateurs se réuniront, en ayant à l'esprit qu'ils devront se déplacer depuis les parkings actifs jusqu'à la scène. **Vous devrez minimiser cette distance.**



Le concert no 6 aura lieu sur cette scène.

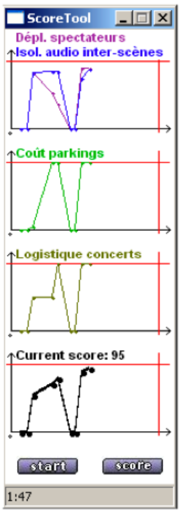
Ne placez pas les scènes sur les routes principales d'accès.

Pour chaque concert, vous devrez poser **un signe sur le signe** représentant la scène choisie (il peut y avoir plusieurs signe concert par scène).

(3)
Vous aurez au maximum <u>45 minutes</u> pour remplir cette tâche. Un outil vous permettra de visualiser votre score intermédiaire tout au long de la tâche.

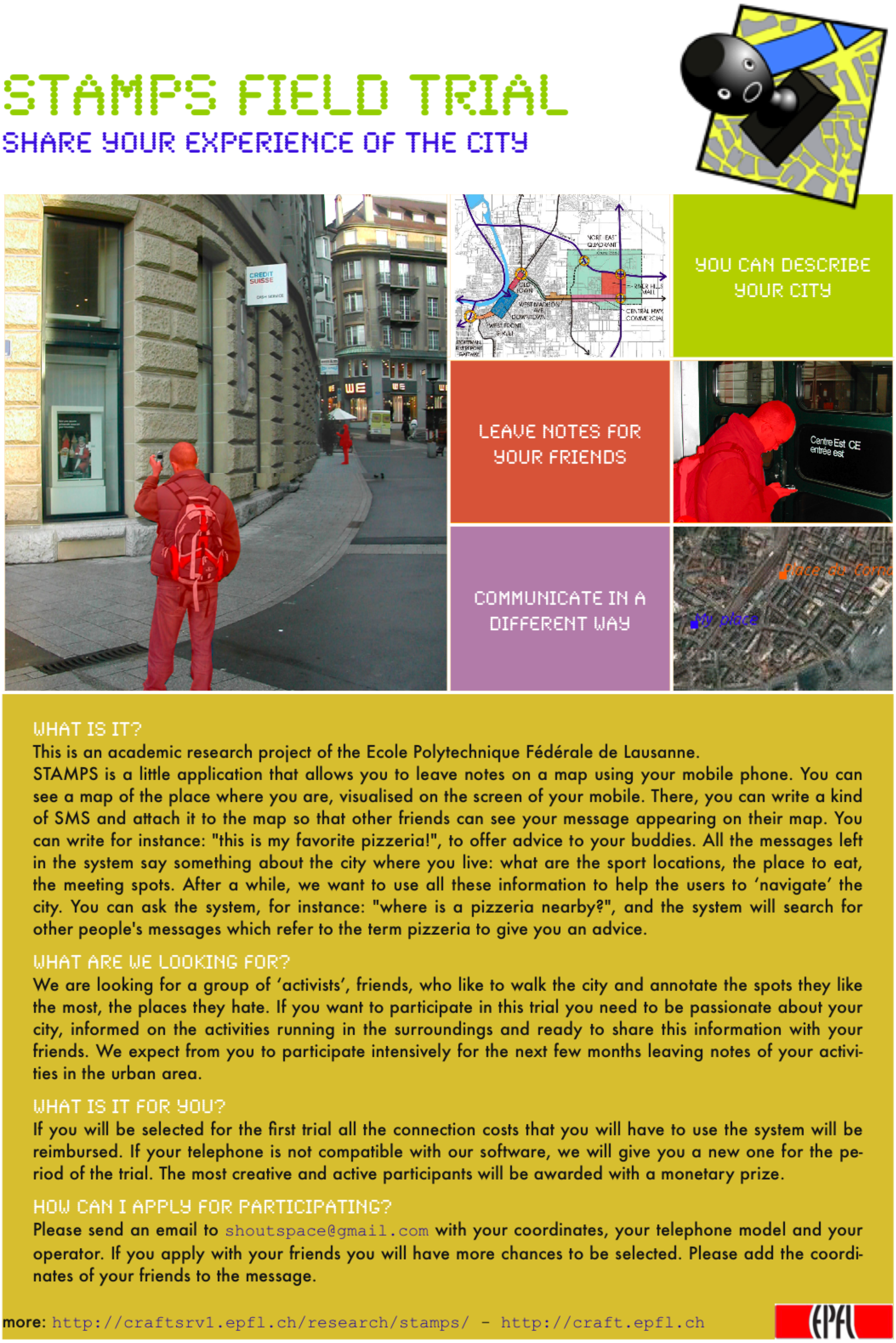
(4)
Le score final dépendra des solutions trouvées pour remplir les quatre missions ci-dessus, le temps utilisé et le nombre d'erreurs de positions des signes entre votre carte et celle de votre partenaire (il n'y a pas de mise-à-jour commune des signes entre les deux ordinateurs).

(5)
Vous devez arriver à une solution commune, dans laquelle les objets (scènes, parkings, concerts) sont placés au même endroit sur les deux cartes. **ATTENTION : les positions des objets sur vos 2 cartes n'est pas synchronisée**



Visualisation du score intermédiaire : le bouton « start » permet de démarrer la tâche, le bouton « score » permet de voir le score intermédiaire.

Figure A-2: Participant instruction sheet used for the 'festival' experiment, page 2


CREDIT SUISSE
Centre Est CE
entrée est
My place

STAMPS FIELD TRIAL

SHARE YOUR EXPERIENCE OF THE CITY

YOU CAN DESCRIBE YOUR CITY

LEAVE NOTES FOR YOUR FRIENDS

COMMUNICATE IN A DIFFERENT WAY

WHAT IS IT?

This is an academic research project of the Ecole Polytechnique Fédérale de Lausanne.
STAMPS is a little application that allows you to leave notes on a map using your mobile phone. You can see a map of the place where you are, visualised on the screen of your mobile. There, you can write a kind of SMS and attach it to the map so that other friends can see your message appearing on their map. You can write for instance: "this is my favorite pizzeria!", to offer advice to your buddies. All the messages left in the system say something about the city where you live: what are the sport locations, the place to eat, the meeting spots. After a while, we want to use all these information to help the users to 'navigate' the city. You can ask the system, for instance: "where is a pizzeria nearby?", and the system will search for other people's messages which refer to the term pizzeria to give you an advice.

WHAT ARE WE LOOKING FOR?

We are looking for a group of 'activists', friends, who like to walk the city and annotate the spots they like the most, the places they hate. If you want to participate in this trial you need to be passionate about your city, informed on the activities running in the surroundings and ready to share this information with your friends. We expect from you to participate intensively for the next few months leaving notes of your activities in the urban area.

WHAT IS IT FOR YOU?

If you will be selected for the first trial all the connection costs that you will have to use the system will be reimbursed. If your telephone is not compatible with our software, we will give you a new one for the period of the trial. The most creative and active participants will be awarded with a monetary prize.

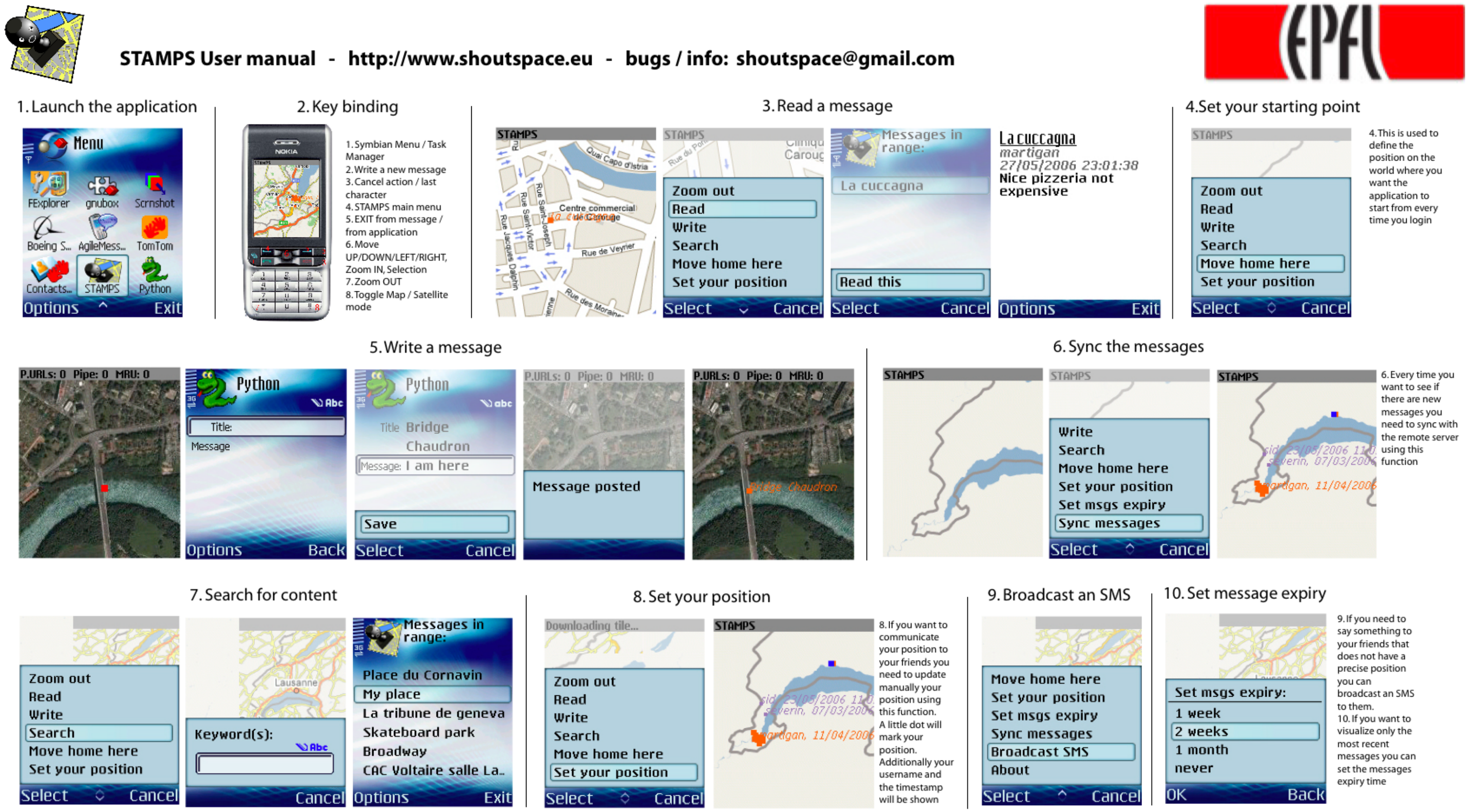
HOW CAN I APPLY FOR PARTICIPATING?

Please send an email to shoutspace@gmail.com with your coordinates, your telephone model and your operator. If you apply with your friends you will have more chances to be selected. Please add the coordinates of your friends to the message.

more: http://craftsrv1.epfl.ch/research/stamps/ - http://craft.epfl.ch

EPFL

Figure A-3: STAMPS field trial recruiting leaflet



STAMPS User manual - http://www.shoutspace.eu - bugs / info: shoutspace@gmail.com

1. Launch the application

2. Key binding

1. Symbian Menu / Task Manager
2. Write a new message
3. Cancel action / last character
4. STAMPS main menu
5. EXIT from message / from application
6. Move UP/DOWN/LEFT/RIGHT, Zoom IN, Selection
7. Zoom OUT
8. Toggle Map / Satellite mode

3. Read a message

La cuccagna
martigan
27/05/2006 23:01:38
Nice pizzeria not expensive

4. Set your starting point

4. This is used to define the position on the world where you want the application to start from every time you login

5. Write a message

6. Sync the messages

6. Every time you want to see if there are new messages you need to sync with the remote server using this function

7. Search for content

8. Set your position

8. If you want to communicate your position to your friends you need to update manually your position using this function. A little dot will mark your position. Additionally your username and the timestamp will be shown

9. Broadcast an SMS

10. Set message expiry

9. If you need to say something to your friends that does not have a precise position you can broadcast an SMS to them.
10. If you want to visualize only the most recent messages you can set the messages expiry time

Figure A-4: STAMPS pocket guide

Appendix B

Log files

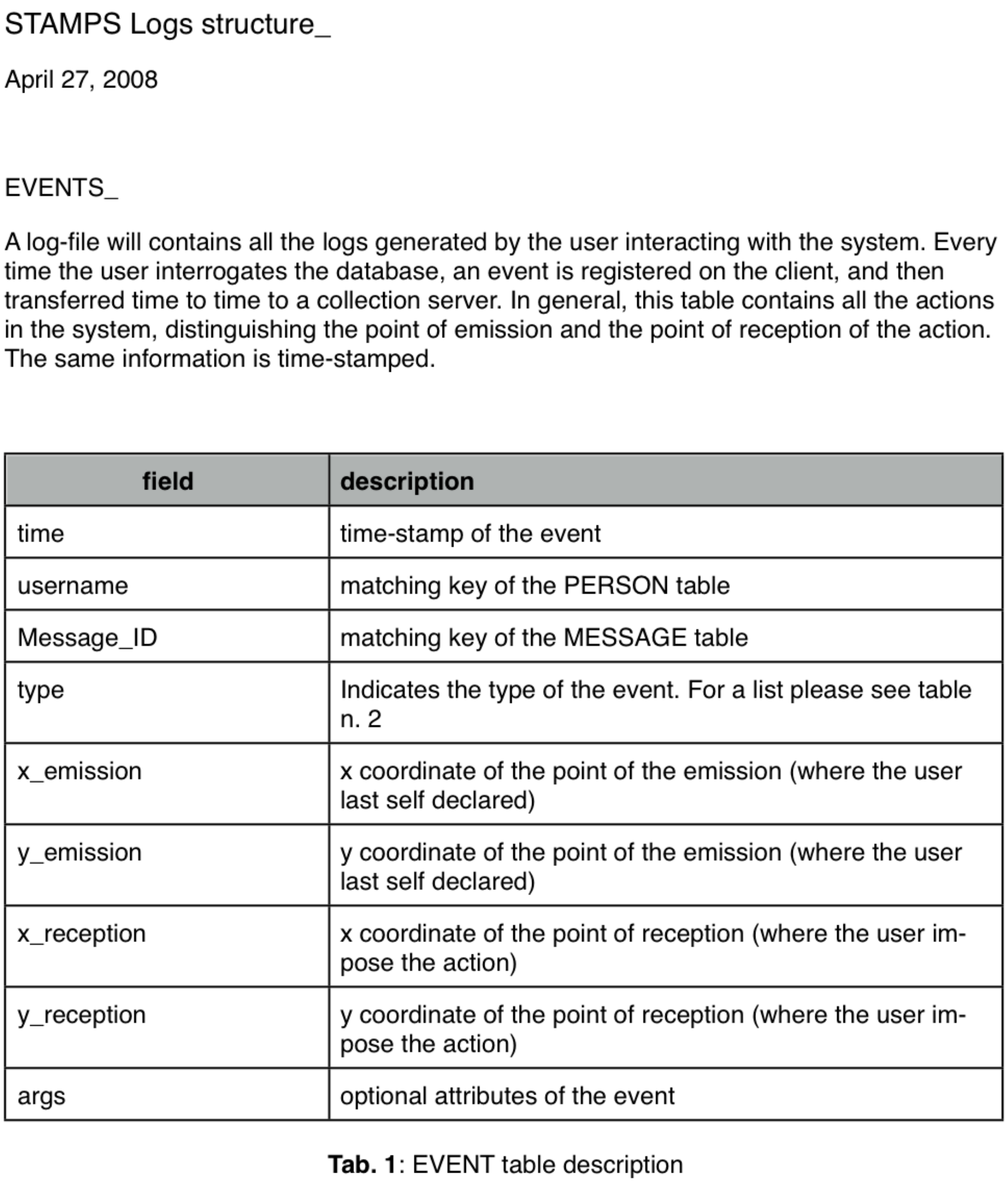
This is an example of the STAMPS log file. Each action in the system is recorded with associated information. The system record the time at which each action occurs (first field), the coordinates of the map currently visualized on the screen and expressed according to the GoogleMap tile system (x, and y on a plane and the zoom level z). Additionally, the Cell-ID of the antenna to which the mobile was connected at the time the action was generated.

STAMPS log, icon_1153775180.0.log:

```
1153775181.0 | 0 | login | 0 | 0 | 0 | 0 | v1.3b | (228, 3, 6001, 11582)
1153775183.0 | 0 | move | 0 | 0 | 128 | 128 | (0, 0, 0) | (228, 3, 6001, 11582)
1153775183.0 | 0 | move | 0 | 0 | 128 | 98 | (0, 0, 0) | (228, 3, 6001, 11582)
1153775217.0 | 0 | zoom_in | 0 | 0 | 1054 | 724 | (4, 3, 3) | (228, 3, 6001, 11582)
1153775220.0 | 0 | zoom_in | 0 | 0 | 2108 | 1448 | (8, 6, 4) | (228, 3, 6001, 11582)
1153775222.0 | 0 | zoom_in | 0 | 0 | 4216 | 2896 | (16, 11, 5) | (228, 3, 6001, 11582)
1153775225.0 | 0 | move | 0 | 0 | 8432 | 5792 | (33, 23, 6) | (228, 3, 6001, 11582)
1153775226.0 | 0 | move | 0 | 0 | 8402 | 5792 | (33, 23, 6) | (228, 3, 6001, 11582)
1153775226.0 | 0 | move | 0 | 0 | 8402 | 5822 | (33, 23, 6) | (228, 3, 6001, 11582)
1153775227.0 | 0 | move | 0 | 0 | 8432 | 5822 | (33, 23, 6) | (228, 3, 6001, 11582)
1153775228.0 | 0 | zoom_in | 0 | 0 | 8462 | 5822 | (33, 23, 6) | (228, 3, 6001, 11582)
1153775230.0 | 0 | zoom_in | 0 | 0 | 16924 | 11644 | (66, 45, 7) | (228, 3, 6001, 11582)
1153775233.0 | 0 | zoom_in | 0 | 0 | 33848 | 23288 | (132, 91, 8) | (228, 3, 6001, 11582)
1153775235.0 | 0 | move | 0 | 0 | 67696 | 46576 | (264, 182, 9) | (228, 3, 6001, 11582)
1153775235.0 | 0 | move | 0 | 0 | 67726 | 46576 | (264, 182, 9) | (228, 3, 6001, 11582)
1153775236.0 | 0 | move | 0 | 0 | 67756 | 46576 | (265, 181, 9) | (228, 3, 6001, 11582)
1153775237.0 | 0 | move | 0 | 0 | 67786 | 46576 | (265, 182, 9) | (228, 3, 6001, 11582)
```

```
1153775292.0 | 0 | mov_hom | 0 | 0 | 0 | 0 | (6.1798095703125, 46.161266875379, 9) | (228, 3, 6001, 11582)
1153775294.0 | 0 | zoom_in | 0 | 0 | 67786 | 46546 | (265, 182, 9) | (228, 3, 6001, 11582)
1153775295.0 | 0 | move | 0 | 0 | 135572 | 93092 | (529, 364, 10) | (228, 3, 6001, 11582)
1153775296.0 | 0 | move | 0 | 0 | 135572 | 93062 | (529, 363, 10) | (228, 3, 6001, 11582)
1153775296.0 | 0 | move | 0 | 0 | 135572 | 93032 | (529, 363, 10) | (228, 3, 6001, 11582)
1153775297.0 | 0 | zoom_in | 0 | 0 | 135542 | 93032 | (529, 363, 10) | (228, 3, 6001, 11582)
1153775301.0 | 0 | zoom_in | 0 | 0 | 271084 | 186064 | (1059, 727, 11) | (228, 3, 6001, 11582)
1153775305.0 | 0 | move | 0 | 0 | 542168 | 372128 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775306.0 | 0 | move | 0 | 0 | 542168 | 372158 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775307.0 | 0 | zoom_in | 0 | 0 | 542168 | 372188 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775326.0 | 0 | zoom_in | 0 | 0 | 1084336 | 744376 | (4236, 2908, 13) | (228, 3, 6001, 11582)
1153775327.0 | 0 | move | 0 | 0 | 2168672 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775330.0 | 0 | mov_hom | 0 | 0 | 0 | 0 | (6.14118576049805, 46.2035963226441, 14) | (228, 3, 6001,
11582)
1153775335.0 | 0 | move | 0 | 0 | 2168702 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775336.0 | 0 | move | 0 | 0 | 2168672 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775350.0 | 0 | move | 0 | 0 | 2168642 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775388.0 | 0 | sync | 0 | 0 | 0 | 0 | [u’65’, u’69’, u’70’, u’77’, u’79’, u’80’, u’86’, u’87’, u’90’, u’91’, u’92’,
u’93’, u’94’, u’95’, u’96’, u’97’, u’98’, u’99’, u’100’, u’101’, u’102’, u’103’, u’104’, u’105’, u’106’, u’107’, u’108’,
u’109’, u’110’, u’113’, u’114’, u’115’, u’116’, u’117’, u’118’, u’119’, u’120’, u’122’, u’123’, u’124’, u’125’, u’126’,
u’127’, u’128’, u’129’, u’130’, u’131’, u’132’, u’133’, u’135’, u’136’, u’137’, u’139’, u’140’, u’141’, u’142’, u’143’,
u’148’, u’150’, u’152’, u’154’, u’155’, u’157’, u’159’, u’160’, u’161’, u’162’, u’163’, u’164’, u’165’, u’171’, u’172’,
u’173’, u’174’, u’175’, u’196’, u’197’, u’198’, u’199’, u’205’, u’216’, u’226’, u’227’, u’228’, u’229’, u’230’, u’231’,
u’232’, u’233’, u’234’, u’235’, u’236’, u’237’, u’238’, u’239’, u’240’, u’244’, u’245’, u’248’, u’249’, u’250’, u’251’,
u’268’, u’262’, u’263’, u’264’, u’265’, u’266’, u’267’, u’269’, u’270’] | (228, 3, 6001, 11582)
1153775389.0 | 0 | move | 0 | 0 | 2168672 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775390.0 | 0 | move | 0 | 0 | 2168672 | 1488782 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775391.0 | 0 | move | 0 | 0 | 2168672 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775405.0 | 0 | move | 0 | 0 | 2168732 | 1488722 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775411.0 | 0 | zoom_in | 0 | 0 | 2168732 | 1488752 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775413.0 | 0 | move | 0 | 0 | 4337464 | 2977504 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775415.0 | 0 | move | 0 | 0 | 4337464 | 2977474 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775416.0 | 0 | move | 0 | 0 | 4337464 | 2977444 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775424.0 | 160 | read | 0 | 0 | 6.14220499992371 | 46.2054383018736 | | (228, 3, 6001, 11582)
1153775527.0 | 271 | reply | 0 | 0 | 6.14212989807129 | 46.2053791658422 | 160 | (228, 3, 6001, 11582)
1153775549.0 | 0 | move | 0 | 0 | 4337434 | 2977444 | (16943, 11631, 15) | (228, 3, 6001, 11582)
```

```
1153775550.0 | 0 | move | 0 | 0 | 4337434 | 2977474 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775551.0 | 0 | move | 0 | 0 | 4337434 | 2977444 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775552.0 | 0 | move | 0 | 0 | 4337404 | 2977444 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775568.0 | 157 | read | 0 | 0 | 6.14000558853149 | 46.2035819815781 | | (228, 3, 6001, 11582)
1153775699.0 | 272 | reply | 0 | 0 | 6.14028453826904 | 46.2035375062293 | 157 | (228, 3, 6001, 11582)
1153775703.0 | 0 | logout | 0 | 0 | 0 | 0 | | (228, 3, 6001, 11582)
1153775772.0 | 0 | zoom_out | 0 | 0 | 4337404 | 2977474 | (16943, 11631, 15) | (228, 3, 6001, 11582)
1153775774.0 | 0 | move | 0 | 0 | 2168702 | 1488737 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775775.0 | 0 | move | 0 | 0 | 2168702 | 1488767 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775776.0 | 0 | move | 0 | 0 | 2168702 | 1488797 | (8471, 5816, 14) | (228, 3, 6001, 11582)
1153775850.0 | 0 | cancel_write | 0 | 0 | 0 | 0 | | (228, 3, 6001, 11582)
1153775852.0 | 0 | mov_hom | 0 | 0 | 0 | 0 | (6.13861083984375, 46.2009247586817, 14) | (228, 3, 6001,
11582)
1153775857.0 | 0 | move | 0 | 0 | 2168672 | 1488797 | (8471, 5816, 14) | (228, 3, 6001, 11582)
1153775858.0 | 0 | move | 0 | 0 | 2168672 | 1488767 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775859.0 | 0 | move | 0 | 0 | 2168672 | 1488737 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775869.0 | 0 | zoom_out | 0 | 0 | 2168672 | 1488767 | (8471, 5815, 14) | (228, 3, 6001, 11582)
1153775872.0 | 0 | zoom_out | 0 | 0 | 1084336 | 744383 | (4236, 2908, 13) | (228, 3, 6001, 11582)
1153775875.0 | 0 | move | 0 | 0 | 542168 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775877.0 | 0 | move | 0 | 0 | 542138 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775878.0 | 0 | move | 0 | 0 | 542108 | 372191 | (2117, 1454, 12) | (228, 3, 6001, 11582)
1153775879.0 | 0 | move | 0 | 0 | 542078 | 372191 | (2117, 1454, 12) | (228, 3, 6001, 11582)
1153775881.0 | 0 | move | 0 | 0 | 542108 | 372191 | (2117, 1454, 12) | (228, 3, 6001, 11582)
1153775882.0 | 0 | move | 0 | 0 | 542078 | 372191 | (2117, 1454, 12) | (228, 3, 6001, 11582)
1153775883.0 | 0 | move | 0 | 0 | 542108 | 372191 | (2117, 1454, 12) | (228, 3, 6001, 11582)
1153775885.0 | 0 | move | 0 | 0 | 542138 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775886.0 | 0 | move | 0 | 0 | 542168 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775888.0 | 0 | move | 0 | 0 | 542198 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
1153775893.0 | 0 | logout | 0 | 0 | 542348 | 372191 | (2118, 1454, 12) | (228, 3, 6001, 11582)
```



STAMPS Logs structure_

April 27, 2008

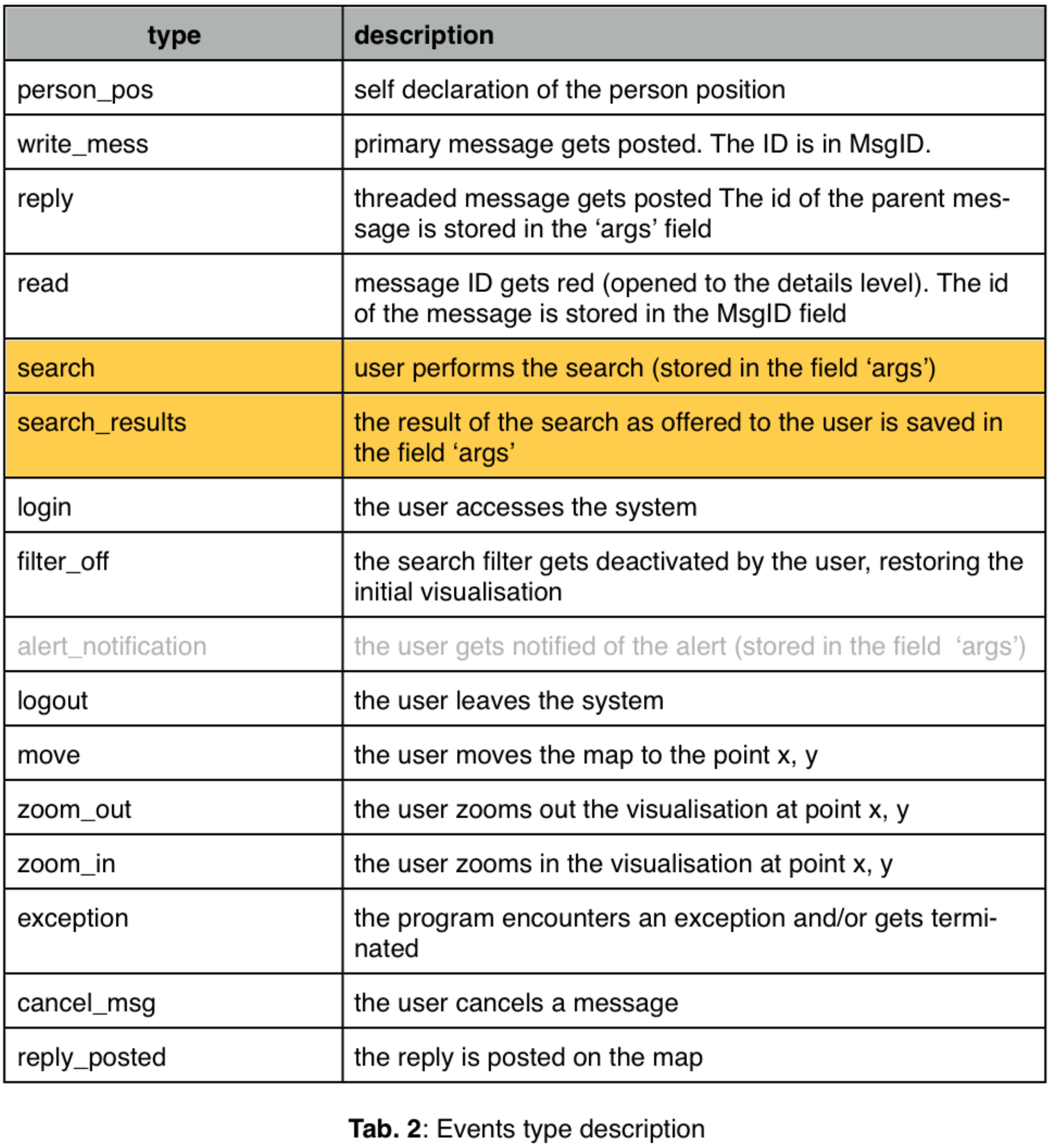
EVENTS_

A log-file will contains all the logs generated by the user interacting with the system. Every time the user interrogates the database, an event is registered on the client, and then transferred time to time to a collection server. In general, this table contains all the actions in the system, distinguishing the point of emission and the point of reception of the action. The same information is time-stamped.

field	description
time	time-stamp of the event
username	matching key of the PERSON table
Message_ID	matching key of the MESSAGE table
type	Indicates the type of the event. For a list please see table n. 2
x_emission	x coordinate of the point of the emission (where the user last self declared)
y_emission	y coordinate of the point of the emission (where the user last self declared)
x_reception	x coordinate of the point of reception (where the user impose the action)
y_reception	y coordinate of the point of reception (where the user impose the action)
args	optional attributes of the event

Tab. 1: EVENT table description

Figure B-1: STAMPS log file structure and action categories, part 1



type	description
person_pos	self declaration of the person position
write_mess	primary message gets posted. The ID is in MsgID.
reply	threaded message gets posted The id of the parent message is stored in the ‘args’ field
read	message ID gets red (opened to the details level). The id of the message is stored in the MsgID field
search	user performs the search (stored in the field ‘args’)
search_results	the result of the search as offered to the user is saved in the field ‘args’
login	the user accesses the system
filter_off	the search filter gets deactivated by the user, restoring the initial visualisation
alert_notification	the user gets notified of the alert (stored in the field ‘args’)
logout	the user leaves the system
move	the user moves the map to the point x, y
zoom_out	the user zooms out the visualisation at point x, y
zoom_in	the user zooms in the visualisation at point x, y
exception	the program encounters an exception and/or gets terminated
cancel_msg	the user cancels a message
reply_posted	the reply is posted on the map

Tab. 2: Events type description

Figure B-2: STAMPS log file structure and action categories, part 2

Below is an excerpt of the log file generated by the Feedback Tool (FT) used during the ‘festival’ experiment. Every time an icon was moved over the map its position was recorded into a separate log file, which was analyzed by the FT every time one of the participants pressed the SCORE button. The x,y screen coordinates were translated into the polygones (here marked as ‘zones’) on top of which the icons were dropped.

Feedback Tool log, feedbackLOG_2006_12_18-15_c.log, experiment 14, MSN condition.

```
15:51:41.89 | info | feedbacktoolopened
16:43:11.13 | start | user pushed start to initialize task | starttime=1166456591.13
17:05:39.03 | info | ID=0 | pushing score to 128.178.88.236
17:05:39.03 | score | ID=0 | TOTAL SCORE=81.8313148227 | walk distance=80.2462668201 |
stages distance=80.4123258039 | planning=100.0 | parking price=66.6666666667
17:05:39.03 | zones | ID=0 | parkings=[’P435’, ’P175’, ’P180’, ’P185’] | stages=[’#space_hn9’,
’#space_hn8’, ’#space_hn6’] | planning=[’#space_hn9’, ’#space_hn6’, ’#space_hn8’, ’#space_hn9’,
’#space_hn6’, ’#space_hn8’]
17:05:39.52 | mistakes | ID=0 | parking=0 | stage=0 | planning=0
17:08:33.23 | info | ID=1 | pushing score to 128.178.88.236
17:08:33.23 | score | ID=1 | TOTAL SCORE=80.1564131277 | walk distance=73.54666004 | stages
distance=80.4123258039 | planning=100.0 | parking price=66.6666666667
17:08:33.23 | zones | ID=1 | parkings=[’P435’, ’P185’, ’P175’, ’P180’] | stages=[’#space_hn9’,
’#space_hn8’, ’#space_hn6’] | planning=[’#space_hn6’, ’#space_hn8’, ’#space_hn9’, ’#space_hn6’,
’#space_hn8’, ’#space_hn9’]
17:08:33.52 | mistakes | ID=1 | parking=0 | stage=0 | planning=0
17:15:27.73 | info | ID=2 | pushing score to 128.178.88.236
17:15:27.73 | score | ID=2 | TOTAL SCORE=80.5079800433 | walk distance=74.9027477206 |
stages distance=80.4625057861 | planning=100.0 | parking price=66.6666666667
17:15:27.73 | zones | ID=2 | parkings=[’P435’, ’P175’, ’P185’, ’P180’] | stages=[’#space_hn9’,
’#space_hn8’, ’#space_n13’] | planning=[’#space_hn8’, ’#space_hn9’, ’#space_n13’, ’#space_hn8’,
’#space_hn9’, ’#space_n13’]
17:15:28.13 | mistakes | ID=2 | parking=0 | stage=0 | planning=0
17:20:44.81 | info | ID=3 | pushing score to 128.178.88.236
17:20:44.83 | score | ID=3 | TOTAL SCORE=80.5079800433 | walk distance=74.9027477206 |
stages distance=80.4625057861 | planning=100.0 | parking price=66.6666666667
17:20:44.83 | zones | ID=3 | parkings=[’P435’, ’P175’, ’P180’, ’P185’] | stages=[’#space_n13’,
’#space_hn9’, ’#space_hn8’] | planning=[’#space_hn8’, ’#space_hn9’, ’#space_n13’, ’#space_hn8’,
’#space_hn9’, ’#space_n13’]
```

```
17:20:45.05 | mistakes | ID=3 | parking=0 | stage=0 | planning=0
17:25:56.27 | info | ID=4 | pushing score to 128.178.88.236
17:25:56.27 | score | ID=4 | TOTAL SCORE=82.4170265315 | walk distance=72.24501371 | stages
distance=90.7564257492 | planning=100.0 | parking price=66.6666666667
17:25:56.27 | zones | ID=4 | parkings=[’P185’, ’P180’, ’P175’, ’P435’] | stages=[’#space_hn9’,
’#space_n13’, ’#space_n37’] | planning=[’#space_n37’, ’#space_hn9’, ’#space_n13’, ’#space_n37’,
’#space_hn9’, ’#space_n13’]
17:25:56.64 | mistakes | ID=4 | parking=0 | stage=0 | planning=0
17:28:09.66 | info | user pushed start but already started
17:28:10.67 | timeover | 1166459290.67
17:46:03.25 | info | feedbacktoolclosed
```

Appendix C

Semantic map of EPFL

Excerpt of the semantic map used to parse the messages extracted from the communication corpus of the ‘festival’ experiment. Each linguistic expression is matched with a polygon identified by the symbol ‘#’. This is a Python dictionary that was directly called by the parsing script.

SEMANTIC MAP EXCERPT:

semanticmap = {′450′:′#P450′, ′195′:′#P195′, ′120′: ′#P120′, ′185′: ′#P185′, ′30′:′#P30′, "180":′#P180′, ′200′:′#P200′, ′250′:′#P250′, ′435′: ′#P435′, ′175′:′#P175′, ′45′:′#P45′, ′210′:′#P210′, ′190′:′#P190′, ′175,′: ′#P175′, "l′esplanade":′#space_n27′, "esplanade": ′#space_n27′, ′pelouse′:′#space_n34′, ′rond point′:′#e_50-43′, ′péninsule′: ′#space_hn6′, "l′espace à gauche du 185?": "#space_n37", ′la péninsule du haut′: ′#space_n13′, ′le grand parking′: ′#P450′, "l′espèce de carré": ′#space_P190′, ′la grande route′: ′#e_56-55′, ′triaudes′: ′#space_P190′, ′PSE′: ′#building_PSE-C′, ′la tente du forum′: ′#space_n34′, ′le champs au sud du 450′: ′#space_P450′, ′batiment de phisique′: ′#building_PH′, ′cp′: ′#building_amphimax′, ′flèche à là droite de la carte′: ′#space_hn9′, ′batiment de farma′: ′#n30′, ′SG′: ′#building_SG′, ′sg′: ′#building_SG′, ′Bm′: ′#building_BM′, ′bm′: ′#building_BM′, ′BM′: ′#building_BM′, ′sous la poste′: ′#space_n37′, ′la sortie du TSOl′: ′#space_hn11′, ′gm′: ′#building_GR′, ′GM′: ′#building_GR′, ′la route cantonal′:′#e_31-2′, ′poly dome′:′#building_PO′, ′TSOL′: ′#n6′, ′polydome′: ′#building_PO′, "triangle rose ′pointe vers le bas′,": ′#space_n39′, ′croisement qui est à la verticale sous le parking à 210′: ′#n14′, ′le croisement qui demare de nuilpart′: ′#n14′, "croisement qui demarre de nulpart,":′#n14′, ′fin de segment′:′#n14′, ′du trapeze rose′:′#space_hn7′, ′le dragon′:′#e_51-43′, ′MT′: ′#building_MA′, ′Agepoly′: ′#building_ME1′, ′agepoly′:′#building_ME1′, ′batiment en H′:′#building_PA′, ′diagonale′: ′#e_33-26′, ′la bichi′: ′#building_BC′, ′bc′:′#building_BC′, ′chimie′:′#building_chimie′, ′H retourné′:′#building_BSP′, ′batiment carré′: ′#building_amphimax′, ′là tente du forum′:′#space_n34′, ′petit coin vert′: ′#space_n39′, ′pointe vers le bas′: ′#house_zone2′, ′MX′: ′#building_MX′, ′le petit coin en bas tout à gauche′: ′#space_n58′, ′les deux parkings en bas ? gauche′: ′#space_hn4′, "le gros de l′esplanade":′#P450′, ′le gros en haut′:

’#P435’, ’le grand espace vert tout en bas,’: ’#space_hn5_2’, ’fen msn collée au carré noir’:’#n52’, ’bichi’: ’#building_CE’, ’MT’: ’#building_BM’, ’sur le dragon...’: ’#space_n34’, ’vers le dragon...’:’#space_n34’, "l’Agepoly": ’#space_n27’, ’batochimie’: ’#building_chimie’, ’du tu’: ’#building_chimie’, ’le cube du park 435’:’#building_amphimax’, ’cube’: ’#building_amphimax’, ’le grand amphi’: ’#building_amphimax’, ’parking GC’:’#space_hn11’, ’GC’:’#building_GC’, ’parc scientifi’: ’#building_PSE-C’, ’h unil’: ’#building_BSP’, ’h en bas’: ’#building_PA’, ’grand espace vert en dessous’: ’#space_n56_2’, ’gro en haut’: ’#P435’, ’un chemin en dessous du H.’: ’#e_23-18’, "le grand de l’esplanade": ’#P450’, ’le grand tout en haut’: ’#P435’, ’pharmacie’:’#n30’, ’un chemin qui le contourne’:’#n30’, ’au T’: ’#building_chimie’, "l’oiseau giratoire": ’#space_n34’, "oiseau":’#space_n34’, ’le batiment H au dessus du park’: ’#building_PA’, ’entre le rouge le chemin et le parking’: ’#space_hn7’, ’un H renvers??’:’#building_BSP’, ’hau à gauche dela fenêtre msn’: ’#n52’, ’entre le batiment en forme de H retourné et le batiment carré’: ’#space_n18_2’}

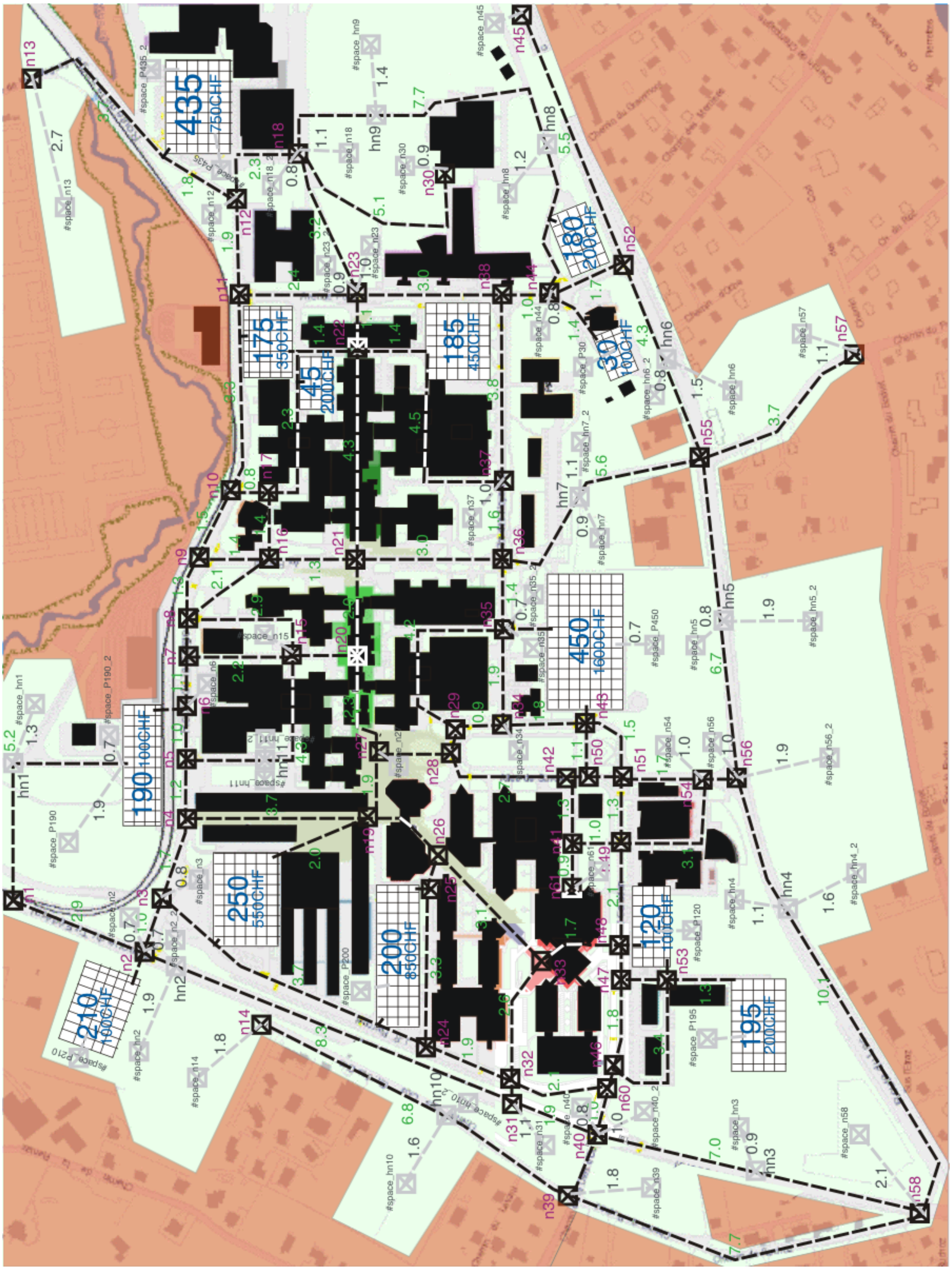


Figure C-1: EPFL campus map with labels of the invisible polygones

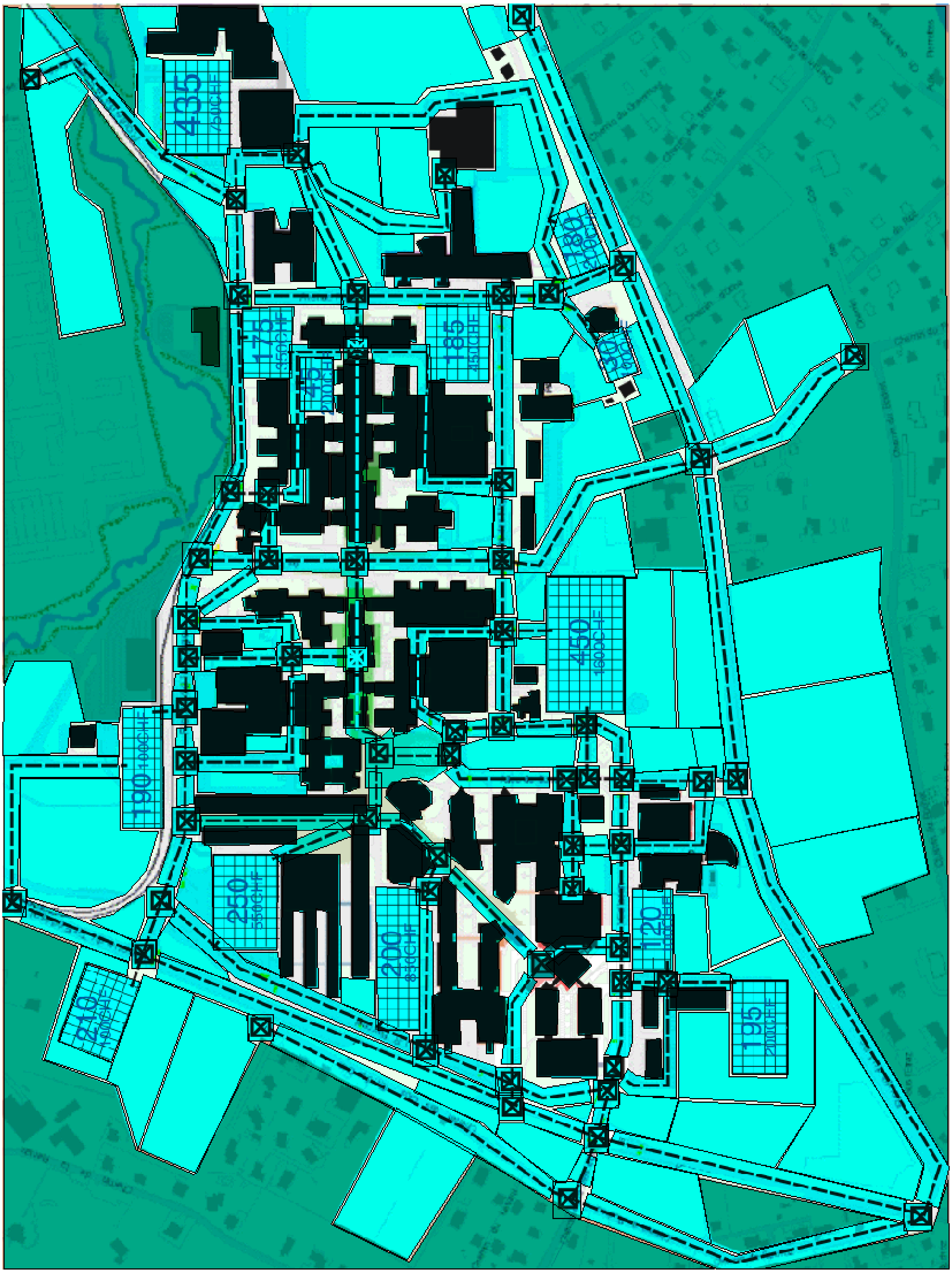
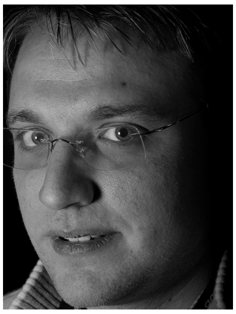


Figure C-2: EPFL campus map with the invisible polygones

Curriculum Vitæ



Mauro Cherubini

CRAFT, École Polytechnique Fédérale de
Lausanne (EPFL), Ecublens, Station 1, CH-1015,
Lausanne, Switzerland
e-mail: martigan@gmail.com
blog: http://www.i-cherubini.it/mauro/blog/
mobile: +39-388.36.36.356
skype: MauroCher

Short biography

Date and place of birth: 7th of February 1977 in Albano Laziale (Rome) – Italy.

After a year in the department of Physics, Mauro Cherubini obtained a degree in Educational Studies from the third University of Rome, RomaTre, in 2001, with a thesis on the Usability of the Children's Internet Sites. He worked for two and a half years as a research assistant at the Media Lab Europe, in Ireland, with several study visits at the MIT Media Lab in Boston. While in Ireland, he obtained a Master of Arts by Research from St. Patrick's College, Dublin City University, in 2004, with a thesis on Microworlds for Ecology Explorations. In 2004, He joined the CRAFT laboratory, at École Polytechnique Fédérale de Lausanne (EPFL).

Selected publications

M. Cherubini, H. Gash and T.J.J. McCloughlin The Digital Seed: An interactive toy for investigating plant growth and the generalized plant life cycle. *Journal of Biological Education*, Institute of Biology Press, London, 2008 (in press)

M. Cherubini, M.-A. Nüssli, and P. Dillenbourg. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the International Symposium on Eye Tracking Research & Applications* (ETRA2008) (Savannah, GA, USA, March 26-28 2008), Association for Computing Machinery, ACM Press.

M. Cherubini, and P. Dillenbourg. The effects of explicit referencing in distance problem solving over shared maps. In *GROUP '07: ACM 2007 International Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), Association for Computing Machinery, pp. 331-340.

M. Cherubini, G. Venolia, R. deLine and A. J. Ko. Let's Go to the Whiteboard: How and Why Software Developers Use Drawings. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI2007) (San Jose, CA, USA, April 28, May 3 2007), ACM Press, pp. 557-566.

Errata corrigenda

Throughout the text I use the expression “significative”, which should be replaced with “significant”. In chapter 8, I sometime used the expression “experiments” to refer to trials.

Last update: 20th of June, 2008.

This thesis was typeset using LaTeX 2$_\varepsilon$with the help of TeXShop on a MacBook Pro.
Bibliographical references were handled by BibTeX. The font used is TeX Gre Pagella.