The Use of Audio and Video in Synchronous Computer-Supported Collaborative Writing

NIENKE NIJENHUIS

Master’s Degree Project
Stockholm, Sweden 2004

TRITA-NA-E04118
The Use of Audio and Video in Synchronous Computer-Supported Collaborative Writing

NIENKE NIJENHUIS

TRITA-NA-E04118

Master’s Thesis in Human Computer Interaction (20 credits) at the School of Computer Science and Engineering, Royal Institute of Technology year 2004
Supervisor at Nada was Kerstin Severinson-Eklundh
Examiner was Kerstin Severinson-Eklundh
Abstract
In this study the use of audio and video and the support of common ground was investigated in the context of the planning stage in the writing process. The research question for this study was: “Compared to the use of an audio communication modality, how is common ground supported by a video communication modality when people are planning a document together with the use of a collaborative writing tool?” To study how common ground was supported by an audio and video communication modality, the following definition of common ground was used: For a conversation to be effective the communicators need to understand each other, that means they are noticing and hearing each other, they understand what was said and what was meant. In total 45 participants took part in this study. The task was to make an outline about Sweden. A questionnaire was used to measure the support for common ground, satisfaction with the interaction and satisfaction with the outcome. Task performance was measured with the quality of the outlines. The outlines were reviewed and sorted on quality. The results for this study show that common ground is well supported by both communication modalities, but differences in support for common ground do exist. It is argued that video supports the creation and maintenance of common ground better in complex situations, which demand intensive collaboration, especially for non-native speakers of English.
Preface

When I started my graduation research project I was not familiar with video conferencing, I was not a good writer and I had not seen much of the world. Now, at this moment that I present my master thesis, I have some experience with video conferences, I have learned a great deal about writing – especially how to write a research report – and I have seen something of the world.

I had told Thea van der Geest I would like to work in Sweden on my project. She suggested I should write a letter of application to Kerstin Severinson Eklundh, when I started to prepare myself for my graduation project. From August 2003 to January 2004 I spent my time in Stockholm, where I was welcomed by Kerstin to do my graduation project at the Interaction and Presentation Laboratory (IPLab), which is a part of the Department of numerical analysis and computer science at the Royal Institute of Technology. I cannot tell you how happy I was that I had the opportunity to go to Sweden, it was a dream I had for ages. In this respect I want to thank Thea and Kerstin a lot for the opportunity to make my dream come true. I want to thank you both for your inspiration and support during the process. I appreciate it a lot!

During my stay in Stockholm I met a lot of wonderful, interesting people. One of the people who offered me a lot of help in the beginning of the project was Eva-Lotta Sallnäs. It was very helpful how we discussed the plans for the study. I also had a lot of fun with Chiara Rossitto, who is really heartwarming in the cold, dark winters of Stockholm. I want to thank all the people that helped me at the IPLab and AMT, I had a very good time with you all.

Further, I would like to thank all my participants and the reviewers who took part in my study. Also, the “afstudeerkring” or “graduation circle” has been a great help for me, especially in the end of the writing process. So, thank you very much Anita Steinmann, Bruno van de Laar, Daniëlle te Velthuis, Gerrit Jan Lutkehaus, JanJaap Gelevert and Nina Kamutzki.

I hope you will enjoy reading my master thesis.

Enschede, July 9, 2004

Nienke Nijenhuis
# Table of content

**CHAPTER 1 INTRODUCTION** .............................................................................................................. 1

**CHAPTER 2 THEORETICAL FRAMEWORK** ......................................................................................... 3
  2.1 Single-Author Writing .................................................................................................................. 3
  2.2 Collaborative Writing .................................................................................................................. 5
  2.3 Collaborative writing tools ......................................................................................................... 6
  2.4 Video and audio communication modality ............................................................................. 8
  2.5 Planning stage of the writing process ....................................................................................... 13
  2.6 Common ground ....................................................................................................................... 14
  2.7 Hypotheses ............................................................................................................................. 19

**CHAPTER 3 DESIGN OF THE STUDY** .............................................................................................. 21
  3.1 Choice for research method ...................................................................................................... 21
  3.2 Participants ............................................................................................................................ 21
  3.3 Materials .................................................................................................................................. 21
  3.4 Procedure .................................................................................................................................. 26
  3.5 Task performance .................................................................................................................... 27

**CHAPTER 4 RESULTS** .................................................................................................................... 29
  4.1 Participants ............................................................................................................................ 29
  4.2 Validity and reliability of the questionnaire .......................................................................... 33
  4.3 Grounding and noticing each other ....................................................................................... 33
  4.4 Grounding and hearing each other ......................................................................................... 35
  4.5 Grounding and understanding each other ............................................................................. 36
  4.6 Satisfaction with the interaction ............................................................................................. 40
  4.7 Satisfaction with the outcome ................................................................................................. 44
  4.8 Task performance .................................................................................................................... 46

**CHAPTER 5 CONCLUSIONS AND DISCUSSION** ........................................................................... 49
  5.1 The support of common ground by audio and video ............................................................... 49
  5.2 Conclusions for satisfaction ..................................................................................................... 50
  5.3 Conclusions for task performance ........................................................................................... 51
  5.4 Reflections on the method ....................................................................................................... 51
  5.5 Suggestions for future research ............................................................................................... 52

**REFERENCES** .................................................................................................................................. 54

**APPENDIXES** .................................................................................................................................. 58
  Appendix A: The video and audio equipment .............................................................................. 59
  Appendix B: Instructions for the participants ............................................................................... 60
  Appendix C: An explanation of the use of Aspects and the task description ......................... 61
  Appendix D: The questionnaire ...................................................................................................... 62
  Appendix E: The instructions to sort the outlines ......................................................................... 69
Chapter 1 Introduction

The way we communicate has changed a lot since the development of computers. Important factors leading to this change were for example the improvement of computer usability, the improved work capacity of computers, the integration of media applications into computers, and the development of broadband network connections. These developments in computers and media applications had a large impact on our daily lives as well as our communication in work situations. Nowadays people depend heavily on the use of computers to share and have access to all kinds of data, documents, spreadsheets, presentations, applications etc. Clearly, this holds for virtually every profession.

As discussed, the possibilities of processors and broadband connections have increasingly transformed computers into multi media devices. In the development of computer supported collaborative writing (CSCWriting), computers were first only used as text editors. Later on, computers were used to work on documents by multiple parties together. After that an audio channel was used to support collaboration and finally it has become possible to add a video channel, e.g. a web cam, to support collaboration between writers. The impact of computer technology on the communication and the final, jointly written product is therefore significant (Galegher and Kraut, 1990), however, there are still gaps in our knowledge about CSCWriting. For this reason, the impact of these developments on the processes of collaborative writing is central to this study.

The study reported here was performed at the Interaction and Presentation Laboratory (IPLab), at the Royal Institute of Technology (KTH), in Stockholm, Sweden. The IPLab conducts interdisciplinary research in human-computer interaction since 1985. The main areas of studies within IPLab are: computer supported cooperative work (CSCW), writing and language technology, and human-robot interaction. Within these fields there is significant attention paid to CSCWriting (Severinson Eklundh, 2002). Research in this area started around 1990 and can be divided in roughly two topics: research on collaborative writing tools as the Domain Help System (DHS) and Col·laboració (see for example Rodríguez, 1999; Rodríguez, Severinson Eklundh, and Gustafsson, 2002; Rodríguez, 2003), and research on the reviewing stage in collaborative writing (see for example Kim and Severinson Eklundh, 1998; Kim, 2001).

Two topics in CSCWriting got relatively little attention. The first of these is the planning stage in the writing process. The planning stage requires intensive collaboration in comparison to composing and reviewing a text. And the second topic is the use of different communication modalities. The design and usability of collaboration tools has received much attention, however, there are only a few reports on the use and effects of specific communication modalities in CSCWriting. Little is known about how to combine different media in ways that are natural for people to use (Isaacs and Tang, 1993). It is important not only to study collaboration tools, but also to study the basic communication modalities used in the tools. Therefore, these two topics in particular will be studied. Gaining more insight in these issues is in the interest of many people involved in collaborations at a distance, for example, writing a scientific article with a colleague in another part of the world, or distance learning.

The topic of this study, the use of communication modalities in CSCWriting, is determined in consultation with the IPLab. The focus of this study is on synchronous communication in the planning stage of the writing process. To support the planning
process, two communication modalities are implemented. The users have either an 
audio channel or an audio plus video channel. The use of these communication 
modalities will be explored using a communication theory on common ground by 
Clark and Brennan (1991). This theory has been chosen because it offers a view on 
how communicators build common ground — that is mutual knowledge, mutual 
beliefs, and mutual assumptions — and how media use influences the processes of 
building common ground. The research question is:

Compared to the use of an audio communication modality, how is common ground 
supported by a video communication modality when people are planning a document 
together with the use of a collaborative writing tool?

This report gives an answer to the research question, on the basis of empirical 
research, done between August 2003 and January 2004. In chapter 2 the concepts 
used in the research question will be explained, which will lead to specific 
hypotheses. Next, in chapter 3 the design for the study will be described. Results of 
this study will be presented in chapter 4. First, the process of common ground will 
come up for discussion, further, results for the audio condition versus the video 
condition will be presented and discussed. Conclusions of this study, as well as the 
discussion for this study, can be found in chapter 5.
Chapter 2 Theoretical framework

The most important concepts used in the research question form the starting point of this chapter. As mentioned in chapter 1, the research question for this study will be: Compared to the use of an audio communication modality, how is common ground supported by a video communication modality when people are planning a document together with the use of a collaborative writing tool?

The context for this study is the act of writing. First the processes involved in single author writing will be presented. Then the processes involved in collaborative writing will be discussed. Furthermore, collaborative writing tools and the audio and video communication modalities are discussed. After that, the planning stage in the writing process and the theory of common ground by Clark and Brennan will be discussed. Thereafter, the choice for the communication modalities in this study will be explained. To conclude the chapter, the hypotheses will be presented.

2.1 Single-Author Writing

Over the last three decades the study of writing has moved from a concern with the written product to descriptive accounts of the writing process (Sharples, Goodlet, Beck, Wood, Easterbrook, and Plowman, 1993). Flower and Hayes (1981) studied the writing process as a problem solving process in the mid-1970s. As a result of their research they developed a model of cognitive processes involved in a writing task. The model identified three main component processes (Flower and Hayes, 1981, p. 369):

- Planning: generating ideas, gathering relevant information to the task, organizing information, and setting goals
- Translation: the turning of plans and ideas into text to meet the goals
- Reviewing: a combination of evaluating the text and editing either the text itself or the ideas and goals

This model is consistent with other descriptions of components or stages in the writing process. Baecker, Nastos, Posner, and Mawby (1993), Faigley, Cherry, Jolliffe, and Skinner (1985), Neuwirth, Kauper, Chandhok, and Morris (1990), Srinivasan Rao, McLeod, and Beard (1996) distinguish between:

- Planning and generating ideas
- Drafting or composing a document
- Reviewing, revision and completion of the document

Sharples (et al., 1993; 1999) and Tammaro, Mosier, Goodwin, and Spitz (1997) recognize that the writing task is recursive. The act of writing might lead to new ideas, which may lead the writer to reconsider the goals of the writing task. That is to say, the three stages of planning, writing, and revising occur continually throughout the writing process.

2.1.1 Planning

According to Faigley et al. (1985), the first stage in the writing process, planning, is rather similar to problem solving strategies. Goals are set and will help writers to reduce the number of constraints they must work within. These goals are frequently tested and re-evaluated in light of what has been written. Additional goals can be generated throughout the writing process and existing ones can be refined. Goals for the writing task can be described in a plan.

Sharples (1999) defines a (writing) plan as a scheme to accomplish some purpose, which needs to be continually reinterpreted to fit the changing situation in the writing process. The writing plan can serve multiple purposes: content, or ‘what to say’, and
rhetoric, or ‘how to say it’. This leads to the construction of several types of writing plans. Sharples (1999) presents a spectrum of types of plans, from content to rhetoric, e.g. notes and sketches (content), mind maps, rhetorical templates and draft texts (rhetoric). See section 2.5 for a more elaborated description of the planning stage.

2.1.2 Composing
Based on the planning activities, the writer should be able to put the ideas for the writing task into words and write a text. Composing, the second stage in the writing process, entails more than just putting ideas into words and sentences. Composing is creating and structuring language to achieve an effect on the reader. It can be seen as a flow of communication between writer and the potential reader (Sharples, 1999). However, the effects of this flow are limited. The reader can never derive the exact meaning of the writer from the text. Sharples (1999) gives two possible explanations for this; the first is that the personal experiences of people are different for each individual, and the second explanation is that written language is a different type of representational system to mental schemata; a mental schema being ‘a cluster of related concepts and constraints that frame thinking and acting’ (Sharples, 1999, p. 44).

The situation in which a writer finds him/herself at any point during composing activities consists of mental schemas and constraints, the text that has just been written and the physical environment. The writer holds a representation of this situation in short-term memory, which provides a context for creative thinking through knowledge telling and knowledge transforming. These are two writing strategies generally employed by writers. Knowledge telling is a strategy in which the writer generates an initial concept or event and then uses that concept as a cue to probe long-term memory for other associated ideas. The writer retrieves a further idea until the ideas stop flowing or the text reaches something recognized as a conclusion. Knowledge transforming is in a way similar to knowledge telling, except that in this strategy explicit mental spaces for content and rhetoric are formed by the writer. The content space contains the writer’s beliefs about the writing topic, and the rhetoric space contains knowledge about the text, e.g. style, structure, purpose and audience (Sharples, 1999, pp. 22, 23).

The context for creative thinking, which is provided by the situational representation in working memory, gives cues that activate long-term memory, retrieve further ideas, or propositions that meet the constraints of audience and purpose, style and structure. The writer then selects and organizes the propositions and turns them into language that fits the context. Finally, the writer transcribes this language into text.

2.1.3 Revision
The third stage in the writing process, reviewing and revision, is generally seen as the reconsideration of text already produced (Faigley et al., 1985). It often starts with reviewing the text at hand, that is, rereading the written text. Expert writers generally review a text to gain overview of the text as a whole, to reflect on the text, or to keep the potential reader in mind. While reading, the writer might decide which parts need to be revised, for example new ideas might be added, grammar and spelling might be checked, etc. The main reason for expert writers to revise a text is to make the text more appropriate and acceptable to their readers. Reviewing and revision can be separate activities from writing, but can occur while composing as well.

Besides the studies of Flower and Hayes (1981), more cognitive studies on single author writing showed that the writing process is a complex and demanding task. Generating ideas, organizing information, setting goals and producing plans may
take a lot of the writer’s time. Recent studies have moved beyond the examination of cognitive processes in writing tasks (Sharples et al., 1993). More attention has been given to external factors influencing the writing processes. For example: the community of which the author is a member of, the author’s attitudes and approaches towards writing, the interaction with other authors who give a context for the task, and the use of computer systems and software designed to support the writing processes (Sharples et al., 1993).

2.2 Collaborative Writing

Writing is seldom a solely individual activity. It often involves interaction with other authors and persons who provide direction, support and comments. Moreover, documents are regularly written in close collaboration with co-authors. Specifically in work situations, the writing process of an individual is embedded in larger-scale processes (Van der Geest, 1995). In comparison to single-author writing, the complexity of a collaborative writing task is magnified many-fold (Sharples et al., 1993). Writers not only have to translate their thoughts and ideas about a text in written language, but also need to communicate their ideas about and attitudes towards the document to their fellow collaborators. In addition, collaborators must share and discuss their thoughts if the group is to establish a shared understanding of the task, and they must negotiate about the constraints on the task as well as the strategies for carrying out the work (Sharples et al., 1993).

Collaborative writing can take many forms and there is no single activity that can be described as collaborative writing (Sharples et al., 1993). The concept of collaborative writing can be characterized as involving two or more people working together to produce a document. The collaborative writing activities involve stages of writing and stages of communicating. It involves periods of synchronous activity, where the group works together at the same time, and periods of asynchronous activity, where group members work at different times. Furthermore, collaborative writing involves shared artifacts that are subject of the work (Miles, McCarthy, Dix, Harrison, and Monk, 1993). Shared artifacts play an important role in the communication between writers, as is elaborated in section 2.2.1.

According to Sharples et al. (1993) central issues in collaborative writing are:

- Task issues, such as partitioning and coordination of work
- Group issues
- Communication issues
- External representation issues, that is, issues in the representation of ideas and constraints for the writing task

This study will focus on communication issues, in particular building common ground in conversations using video and audio. In the next section communication issues around artifact sharing in collaborative writing will be discussed, leading to section 2.3 in which collaborative writing tools will be discussed.

2.2.1 Shared artifacts

In collaborative writing the (writing) task is shared amongst co-authors. This shared task involves some artifacts that are the subject of the work. These artifacts may be entirely conceptual, e.g. a joint decision, or be physical, such as a shared document. The artifacts may or may not be part of a computer system (Miles et al., 1993). The interaction with an artifact between two or more authors can be synchronous or asynchronous depending on the chosen writing strategy (Posner and Baecker, 1992). Writing strategies include:

- Single person writing, that is when one person writes the document based on discussions with other group members.
- Scribe, that is when one individual takes the role of writing down the group’s thoughts in a group meeting.
- Separate writers, that is when the document is broken up in parts with each group member writing and being responsible for a different part.
- Joint writing, that is when the document is being written together in a group, in which group members can decide together on the exact wording and sentence structure used in the text.

The last strategy was used in this study, in which participants worked in pairs on a writing task. Posner and Baecker (1992) suggest that early on in a collaborative writing task, groups can write jointly to produce an outline of their ideas, at which point opinions are still forming. Later on in the task, the individuals’ ideas are better defined and more difficult to integrate.

Besides the interaction with an artifact, collaborators communicate indirectly through an artifact. For example, one can think of annotations in a written document, characteristics of the physical appearance of a document, and writing style (Miles et al., 1993). In addition to interacting with and through the artifact, writers will use various means to refer to particular artifacts. To successfully refer to a particular artifact Clark and Brennan (1991) describe several means. In face-to-face communication one way of referring to artifacts is the ‘indicative gesture’, which highlights the importance of pointing, looking, and touching as a means of grounding references. Mutually establishing references to artifacts plays an important role in building common ground – that is mutual knowledge, mutual beliefs, and mutual assumptions (see section 2.6), because many conversations focus on artifacts and their identities.

2.3 Collaborative writing tools

One particular shared artifact in this study is the shared editor, which the participants use to carry out the writing task. A shared editor is one of many types of collaborative writing tools. Basically, a shared editor is a word processor which shares the interface amongst a number of computers. This way, the collaborators will have a similar view of the shared document. The word processor was originally designed to support the writing process from a narrow point of view, that is to say, exclusively to support the writer in writing down and revising text. “The facilities added onto word processors were designed to support separate sub-processes: an outliner to support planning, a thesaurus to help finding the appropriate and varied wording, a spelling and grammar checker to help the writer editing the text” (Van der Geest, 1995, pp. 7, 8). In addition, collaborative writing tools were designed from a collaborative work point of view, in particular to support communication and information sharing between the collaborators.

In the field of CSCWriting much research has been done on the design and development of web based collaborative writing tools (Noël and Robert, 2003). Several web-based collaborative writing tools are described and compared in an article by Noël and Robert (2003). The authors describe the following advantages and disadvantages of web based writing tools in comparison to local area network (LAN) and wide area network (WAN) based tools. Many people already have access to the Internet, making it more likely to find co-writers. Due to the accessibility of the web, web based tools are available to users of all types of computers. And, users of web based tools do not need to install and maintain new programs on their computer. The disadvantages of web based collaborative writing tools are: the difficulty of implementing synchronous work on the web, the unstableness of the Internet (e.g. the connection might get interrupted at any time, web pages might not be accessible,
etc.), and the web’s user interface is much less evolved than that found in a graphical user interface environment (e.g. Mac OS, Windows).

The web based collaborative writing tools described in Noël and Robert (2003), show many deficiencies in supporting communication in the writing processes. None of the tools described offers a specific communication function, it seems as if these tools were designed to support those activities that require the least interaction between collaborative writers (writing, editing, and reviewing). Synchronous communication is not supported either. According to the authors this would be too complex to incorporate into the tool. What the collaborative writing tools do support is amongst others:

- Uploading and downloading of documents used in the writing task, from server to work station and vice versa.
- Email notification to the users of the tools, for example when changes have been made.
- Annotating the product of the writing task.
- Social awareness, which means having an understanding of the activities of other group members, which provides a context for one’s own activities (Posner and Baecker, 1992).
- Locking of text as a form of coordination, to protect the text from being changed, during work in e.g. synchronous activities (Posner and Baecker, 1992).
- Uncomplicated writing activities, such as researching, writing, editing, and reviewing.
- Social roles; Posner and Baecker (1992) distinguish between ‘writer’, ‘consultant’, ‘editor’, and ‘reviewer’ roles in the writing process. Misunderstanding of the social roles group members play in a writing task may lead to conflicts within a group.
- Protection of the work, e.g. protect work from getting erased accidentally.

However, the authors conclude that different tools support different aspects. Not one tool supports all aspects. And when certain aspects are supported, different tools support them often in different ways.

2.3.1 The choice for the Aspects tool
Overall, the discussion of web based tools in Noël and Robert (2003) puts forward that synchronous communication, as well as the planning stage are not supported sufficiently by web based collaborative writing tools. This led to the choice of using the LAN based shared editor Aspects, which runs on MAC OS.

Aspects, designed by Group Technologies, is a collaborative conferencing system that runs on networked computers and provides writing, drawing, and painting tools. The writing tool in Aspects has two components. First, the Collaborative Word Processor allows sharing a document across the screens of the computers in the same network. Each group member can view and edit the text together, and see each other's changes as they happen. Second, the Chat Box, allows group members conduct an electronic discussion by posting comments. The participants in this study did not use this chat box, but used either an audio or video modality to communicate. In the next section the audio and video communication modalities will be presented.

The shared editor supports according to Posner and Baecker (1992):

- Preservation of identities within the group of authors; when identities of individual writers are preserved communication is more effective when feedback is send to the author of a segment.
- Synchronous communication
- Explicit social roles, like writer, consultant, editor and reviewer
- Brainstorming
- Planning
transitions between activities: brainstorming, planning, writing and editing
- Accessing relevant information
- Version control, that is, the accessibility of the information about which changes were made in a document, who made these changes, and when were they made.
- Synchronous and sequential document access
- Write and read-only access methods
- Working on separate document segments
- Single-author writing as well as collaborative writing (any number of writers could use Aspects)
- A synchronous and asynchronous writing approach (Posner and Baecker, 1992).

These aspects, such as making social roles explicit in the tool, are not imposed on the user of this tool, but can be chosen for, when starting collaboration.

2.4 Video and audio communication modality

Apart from shared artifacts such as collaborative writing tools, writing activities are influenced by the communication modalities used in these activities. In this study the influences of the audio communication modality will be compared to the video communication modality on support for common ground. First the choice for the modalities will be explained. Thereafter, earlier research on the communication modalities will be discussed. Finally, the section ends with the important issues for this study.

2.4.1 The choice for the audio and video communication modality

Previous research on the use of different media in CSCW in general, supports the hypothesis that the text based medium is poor in comparison to the audio and video medium (Bos, Olson, Gergle, Olson, and Wright, 2002; Sallnäs, 2002). Also, previous research supports the hypothesis that face-to-face communication is the richest medium, in comparison to audio and video modalities (Olson, Olson, and Meader, 1995; Sallnäs, 2002). In addition, it has been suggested that adding audio to any text-based communication modality will enhance the interaction. However, adding a video channel to audio communication did not directly improve the interaction (Nowaczyk, Thomas, and White, 1991). Therefore, audio and video have been chosen to study, also since the choice of the medium in CSCWWriting is important for the development of a collaborative writing tool (Andersson, 2003).

The main feature that distinguishes video from audio is visibility (Clark and Brennan, 1991). Visual information is used to coordinate among actions and talk and body language forms a context for what is said. It helps people maintain up-to-date mental models or situational awareness of others’ activities. This way, people can coordinate what to say next and when to say it. Visual information can also help people communicate about the task, a form of meta-communication, to develop mutual understanding between the conversational partners. Moreover, visual information plays an important role in establishing mutual understanding, or common ground.

Bradner and Mark (2001) and Fussell, Setlock, and Kraut (2003) suggest that there are four sources of visual information about the other conversational partner: partner’s head and face (facial expressions), partner’s body and actions (gestures, eye-gaze, body position), the (shared) artifacts and the work environment. These sources of visual information are linked to five social key aspects: monitoring task status, monitoring partner’s actions, identifying what one’s partner is attending to, formulating messages and monitoring partner’s comprehension. These social key
aspects are linked to the concepts of common ground which are central to this study, see section 2.6.

2.4.2 Previous research on communication modalities

The video communication modality is often studied in the combination with face to face communication and the audio communication modality. In the following section, several studies will be discussed which studied these modalities.

Isaacs and Tang (1993) studied the implications for user interface design when using video for remote collaboration. The authors argue that the effects of video are more likely to be visible when studying the process (instead of the product) of interactions, particularly among people who know each other and are accomplishing real work, as opposed to laboratory experiments. They observed a team of software engineers (n = 5) who were using a desktop video conferencing prototype in real work situations. The use of this video conferencing prototype was compared to the use of an audio-only communication modality and face-to-face communications.

The intention of the authors was to describe the benefit of video in remote conversations over audio alone, and to point out how video interactions fall short of, and in some ways offer advantages over, face-to-face interactions (Isaacs and Tang, 1993, p. 199). Six interactions were studied, but information about the method to analyze these interactions and which variables were measured is missing. In the video conferencing condition the authors analyzed the facial expressions, gestures, and body language. Their hypothesis was that conveying the information over the audio channel would take more effort. However, this hypothesis was not completely supported with data from the observations.

The authors found that video in e.g. managing the mechanics of conversations had a more positive influence compared to the audio condition. Furthermore, compared to audio-only, the video modality added or improved the ability to show understanding. In addition, using video improved forecasting responses, giving non-verbal information, enhancing verbal descriptions, managing pauses and expressing attitudes. Video, is their conclusion, might be especially useful for handling conflict and other interaction-intense activities. Another finding in this study was that the authors expect to see long term effects rather than short term effects, such as more frequent video collaborations, due to the participants’ satisfaction with the medium. However, according to the authors, the advantages of video depend critically on the nearly-instantaneous transmission of audio. To fully enable rich interactions, video should be integrated with other distributed tools. For example, a shared writing tool, that increases the extent and type of shared space in such a way that natural collaborative behavior within those environments is enabled. Although the findings of this study are very interesting, the findings of this study are not empirically supported. The conclusions are merely based on suggestions from the observations.

Fish, Kraut, Root, and Rice (1993, p. 50) reviewed previous literature consistent with the following hypotheses that video-based communication is helpful in:

1. increasing the spontaneity and frequency of communication
2. supporting social relationships
3. coping with the most complex and equivocal communication problems encountered in work groups, and as a result
4. integrating members into and supporting the work in research and development groups

In addition to their literature overview, the authors describe an experiment to evaluate the adequacy of audio and video conferencing for supporting informal communication in a realistic setting. A four-week field experiment was conducted, in which 23
Volunteers participated, 11 students and their 12 mentors. The participants used an audio-video conferencing prototype in a summer internship program. The audio-video conferencing prototype (hereafter ‘prototype’) was supplemented by three calling methods to encourage spontaneous interactions that might lead to conversation. The first calling method was called ‘Cruise’, which consisted of one or a series of audio and video calls. Individuals and a list of users could be called, as well as a random group of users. The second calling method was called ‘Glance’, which were one second video-only connections to one or a series of other users, these calls could also be made to a random group of users. The third calling method was called ‘Autocruise’, in which the prototype itself initiated calls between users at random times. The intended analogy for the Autocruise was wandering in a corridor and randomly seeing other people with whom one could speak (p. 51). The variables in this experiment were:

- the number of calls made using the prototype
- the length of the calls
- the type of the calls made, according to the three calling methods
- self-report information from debriefing interviews comparing a conversation in the prototype with a face-to-face conversation
- privacy issues discussed during debriefing interviews
- perceptions of the usefulness of nine different media for a variety of communication tasks

The findings for this field experiment show that the prototype was adequate, but only marginally for the support of spontaneous conversations. Conversations in the prototype were seen as more similar to telephone conversations then face-to-face conversations, in spite of the use of video. These findings did not support the expectation of the authors, which was that the use of video in informal conversations would be similar to informal face-to-face conversations. In this context, Campbell (1999) points out the differences between face-to-face conversations and video based conversations in:

- The level of conflict and misunderstanding (this is higher with video)
- The structure of meetings (more structure is needed in video)
- Task focus – the number of interruptions and distractions during meetings (video is more formal)
- Problem complexity and the degree and frequency of group interaction
- Sociability – the degree of social presence supported by the medium (see also media richness theory, section 2.6.2)
- Communication richness – how easily participants are able to develop a shared understanding and consensus.

Olson et al. (1995) conducted a study in which participants were connected by a shared editor and either high quality stereo audio or the same plus high quality video. Thirty-six groups of 3 MBA professionals took part in this study. In these small teams, the synchronous communication was studied and compared with a preceding study on the face-to-face communication and audio communication without the use of the shared editor. The variables were the following: the quality of the product, the participants' satisfaction with the group process, and with the process of task and coordination during the collaboration. Unlike the studies by Isaacs and Tang (1993) and Fish et al. (1993) who studied media in realistic settings, a specific design task was chosen in this study. This task was chosen because it was a representative ill-structured problem solving task that interleaves many sub processes, such as planning, creativity, decision-making and cognitive conflict (Olson et al., 1995, p. 363).
Video appeared to add value in comparison to audio on satisfaction with the group process and with the processes of design and coordination partially, but no significant differences were found for the quality of the products. Based on these results, the authors argue that it is possible to produce work indistinguishable in quality from face-to-face work, when using high quality communication modalities and a shared workspace tool. The audio-only groups were only marginally different from the video groups. Thus, the authors conclude, high quality intellectual group work is possible under distributed conditions.

In a study on comparing touch, text, voice and video in collaborative virtual environments (CVE’s), Sallnäs (2002) argues that using video might be more enjoyable – whereas using audio might be more efficient. Findings on the number of words used per second would suggest that video is a more social medium than audio is. These conclusions were based on an experimental study in which forty pairs of participants took part. Three different communication modalities were investigated: text-chat, audio conference and video conference. The CVE was compared to a control condition implemented on the Web. In the Web environment two communication modalities were used: audio and video. The participants had to perform a decision-making task. The variables measured in this study were: perceived social presence, perceived virtual presence, perceived task performance. Further, objective measures of the interaction were: time to finish the task, frequency of words used, and the number of words per second.

Kraut, Fussell, and Siegel (2003) conducted two experiments, in which the value of head-mounted video was evaluated. This head-mounted video system provided remote helpers with a view of what the worker was looking at and a portion of the surrounding environment. This system was compared with an audio-only condition in experiment 1, and in experiment 2 to a side-by-side condition in which helpers and workers were co-located. In the first experiment 60 undergraduate students participated. The performance of the participants on bicycle repair tasks was compared in three conditions: working alone, working with a helper connected by audio-only, working with a helper connected by a video link. The following hypotheses were tested:

- Performance in the video condition was expected to be better than that in the audio-only condition.
- Helpers would provide better instructions (in content and timing) in the video condition.
- Workers in the audio-only would describe more explicitly what they are doing and the status of the task.

Performance measures were taken from the number of tasks completed, task completion time and repair quality. Real-time observations of the interactions were coded into work quality, and helper and worker communication by using a five-point scale from 1 (poor) to 5 (well). All the interactions were recorded and verbatim transcripts were made for detailed post experimental coding of the communication. Results of the first experiment showed that the hypothesis that participants in the video condition would perform better was not supported. However, helpers did provide better instructions in the video condition, and the workers’ descriptions were significantly less explicit.

In experiment 2 a within-group experimental design was used, in which each pair conducted tasks under all communication conditions. Workers performed in this experiment three repair tasks on a 10-speed bicycle with the assistance of either an expert or a novice helper. Pairs performed one task in each of three media conditions: side-by-side (or face-to-face), audio-video, and audio-only. Twenty-five students took part in this study.
The following hypotheses were tested:

- Performance would be best in the side-by-side condition, intermediate in the video condition and poorest in the audio-only condition.
- Conversational grounding, as indicated by message length, number of conversational turns, and use of deictic expressions (referential expressions), should be easiest in the side-by-side condition and hardest in the audio condition.
- Deixis and pointing gestures would be most frequent when pairs worked side-by-side. Deixis was measured by short-hand expressions and pronouns such as "this one".
- The use of deixis and pointing gestures in the video condition would not be as frequent as in the side-by-side condition, but more frequent than the audio-only condition.

The same measures were taken as in experiment 1. Results of the second experiment provided additional support for the hypothesis that the presence of visual information affects conversational grounding in collaborative physical repair task (hypothesis 2). The visual information that the helper received over the video system influenced the form of pairs’ dialogues (hypothesis 3) but did not improve performance over that in the audio-only condition (hypothesis 1).

This section discussed previous research on influences of audio and video communication modalities in collaborative work. To summarize the section, video was found to add more value than audio to manage the mechanics of conversations, the ability to show understanding, to forecast responses, to give non-verbal information, to enhance verbal descriptions, to manage pauses and to express attitudes (Isaacs and Tang, 1993). However, these findings were not supported by empirical data. Fish et al. (1993) found that video did not support informal conversations in comparison to face-to-face communication. Nevertheless, Olson et al. (1995) concluded that it is possible to produce work indistinguishable in quality from face to face work, when using high quality communication modalities and a shared workspace tool. Sallnäs (2002) found that using video might be more enjoyable – whereas using audio might be more efficient. The findings suggested that video is a more social medium than audio is. The conclusions from the study by Kraut et al. (2003) were that video supports to provide good instructions better than audio, and the authors found evidence that visual information in the video system influenced the form of the pairs’ dialogues. The presence of visual information affected conversational grounding in collaborative physical repair as well. Two issues were repeatedly studied in this field, namely satisfaction and task performance.

Satisfaction and task performance will be discussed in the following sections.

### 2.4.3 Satisfaction

Olsen et al. (1995) studied the participants’ satisfaction with the group process, and with the process of design and coordination during the collaboration. They found that video appeared to add value in comparison to audio on satisfaction with the group process and partly to the processes of design and coordination. To assess satisfaction, the authors constructed a post-session questionnaire that asked the participants to rate their satisfaction with the form of interaction that they used, as well as with the design result, assess the evenness of the participants’ contributions, and identify a leader if one emerged. In a study comparing four communication modalities, Suh (1999) assessed task satisfaction in two constructs: Process satisfaction, which refers to a cognitive state resulting from the group problem-solving process itself, and solution or outcome satisfaction, which refers to the degree of fulfilment that an individual seeks. Decision process and outcome satisfaction were measured by a post-session questionnaire to evaluate task satisfaction. Task satisfaction is generally seen as an important indicator of user acceptance of a
system (Suh, 1999). However, no hypothesis about task satisfaction was supported. For a more detailed discussion of this study see section 2.6.2.

2.4.4 Task performance
The hypothesis for task performance has often been that task performance would be of a higher quality when using video in comparison to audio (Kraut et al., 2003; Sallnäs 2002; Suh, 1999). This was expected because people can use visual information when using video. However, findings show no difference between audio-only and video (Kraut et al., 2003; Olson et al., 1995; Sallnäs 2002; Suh, 1999). It is surprising that, although people are more satisfied collaborating by video than by audio-only, it has not been proved that video contributed anything to the outcome performance of people engaged in variety of different tasks (Veinott, Olson, Olson, and Fu, 1999). Veinott et al. (1999) suggest that the only tasks for which video has been shown to add value to the interaction and outcome are certain kinds of negotiation tasks, where presumably people benefit by being able to read important cues from each other’s faces and adjust their strategies accordingly. Besides that, Isaacs and Tang (1993) expect video to be especially useful for handling interaction-intensive activities, such as planning activities in a collaborative writing task. In the following section the activities in planning will be elaborated.

2.5 Planning stage of the writing process
In paragraph 2.1.1 the purpose of planning was presented, as well as the ‘product’ of planning, a writing plan. In this section the focus will be on the activities in planning. The following activities can be distinguished:

- starting up the planning stage
- defining goals and sub goals for the writing task (Faigley et al., 1985)
- generating ideas: brainstorming about content and taking notes (Sharples, 1999)
- negotiating about the generated ideas (Posner and Baecker, 1992)
- generating content
  - Some of the strategies to generate content can be (Faigley, et al., 1985):
    - knowledge telling and knowledge transforming, see section 2.1.2
    - associative, trial and error strategies to generate ideas for content
  - brainstorming about rhetoric, organization and structure of the writing material (Sharples, 1999)
- discussion and production of a (more elaborated) writing plan (Sharples, 1999)
- testing and reevaluation of goals set
  - Some of the strategies to produce new operative goals from existing text (Faigley et al., 1985):
    - pursuing an interesting feature in the existing text
    - looking for contradiction in or objection to what has been written
    - discovering what message really is important to convey
    - looking for a focus
  - organizational planning: dividing the (future) work (Posner and Baecker, 1992)
    - Specifically for a collaborative writing situation, the writers should decide who will write which section, who will edit the whole document, who has the ‘official’ copy at any time, etc. (Noël and Robert, 2003).

This list is not pretending to be exhaustive or to prescribe a sequence in planning activities. According to Faigley et al. (1985) plans do not always follow a hierarchical sequence, but are often based on opportunism.

As can be seen in the list, the activities involved in planning are often ill-structured. Negotiating, decision making, and cognitive conflict are all important sub processes in the planning stage, especially when planning is a part of a collaborative writing
task, as opposed to single-author writing. Therefore, it is important that the collaborators can achieve an appropriate level of common ground. The planning stage can play a major role in a successful writing task, when collaborators understand each other and share the same basis of communication. In the following section will be described how collaborators can achieve appropriate level of common ground, and how media influences the development of common ground.

2.6 Common ground

Up to now the context of this study and issues in collaborative writing have been presented. In this section the theory that is used to frame this study will be presented. This theory has been chosen because it offers a view on how communicators build common ground. In this section the theory of common ground will be elaborated, the media richness theory will be discussed and previous research on common ground will be discussed. The section ends with the key aspects of common ground for this study.

2.6.1 The theory of common ground

The theory of common ground stems from analyzing everyday spoken language (Clark and Marshall, 1981). As a result, Clark and Brennan (1991) take conversations as the perspective for their theory. Spoken language in conversations often is very ambiguous and only works because the parties actively collaborate to make it work. Therefore, the use of spoken language is efficient because only the information relevant to each individual's separate needs is communicated (Monk, 2003).

Contributions to conversations are generally divided in two stages: the presentation stage, and the acceptance stage. In the first stage A presents utterance $u$ for B to consider. A does so on the assumption that, if B gives evidence $e$ or stronger, A can believe that B understands what A means by $u$. In the latter stage B accepts utterance $u$ by giving evidence $e$ that B believes B understands what A means by $u$. B does so on the assumption that, once A registers that evidence, A will also believe that B understands. It takes both stages for a contribution to a conversation to be complete.

Giving evidence $e$ can be negative when a message or utterance is not understood. Besides giving negative evidence, the three most common positive forms of evidence of understanding an utterance or message are: acknowledgements, or back-channel responses like $uh$, $huh$, $yeah$, relevant next turn, and continued attention. The description of the two stages in contributing to a conversation forms the basis of the theory of common ground. In this process of contributing to a conversation Clark and Brennan (1991, p. 224) identified four states in which a receiver can be after the communicator sent a message: 0. the receiver did not notice the message, 1. the receiver did notice the message, but is not in state 2, 2. the receiver correctly heard the message, but is not in state 3, and 3. the receiver understands what was said and the meaning of the message.

Clark (1996, p. 95) defines common ground as follows: “A proposition $p$ is only common ground if: all the people conversing know $p$; and they all know that they all know $p$; and they all know that they all know that they all know $p$”. Another definition (Clark & Brennan, 1991), based on the four states in which a receiver can be, is: “For a conversation to be effective the communicators need to understand each other, that means they are noticing and hearing each other, they understand what was said and what was meant”. In this definition the importance of giving evidence of understanding is unclear, this is more emphasized in the following definition by Kraut, Gergle, and Fussell (2002, p. 33): “Throughout a conversation participants are
mutually assessing what each other knows at any moment, and then using this knowledge to form their subsequent utterances. People are obligated both to assess and give off cues that indicate their understanding”.

As the meaning of the communicator can never be fully understood by the receiver, people use a criterion or basic level on which the receiver must have at least understanding. The criterion communicators try to reach in communication (Clark and Brennan, 1991, p. 223) is defined as: “The contributor and his or her partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes”. Using media, there are constraints in achieving common ground. These constraints are called ‘costs of grounding’: different features affect the costs for speaker and listener to ground communication. Clark and Brennan (1991) distinguish between the following costs of grounding:

- cost of formulation, which means deciding what to say
- cost of production, which means saying it
- cost of reception, which means hearing it
- cost of understanding, which means understanding it
- cost of start-up, which means starting a conversation
- cost of delay, which means the impact of a delay on grounding
- cost of asynchrony, which means the impact of misordering on grounding
- cost of speaker change or multiple speakers
- cost of display / pointing / graphical input
- cost of errors in production or in understanding
- cost of repairs

Group members balance the perceived costs for formulation, production, reception, understanding, start-up, delay, timing, speaker change, display, errors, and repair. For example in collaborative writing, as more participants join the conversation and the medium must support the work of a whole group, costs and tradeoffs shift. Start-up costs may be greater. Perception costs will increase if a hearer must put effort into identifying who is speaking or writing. Error costs and repair costs will be higher when a group is involved. Any medium that supports cooperative work can be evaluated in terms of the techniques it allows for grounding.

There are three primary sources for common ground: first, people may have common ground prior to an interaction if they are members of the same group or population. Second, people can construct and expand their common ground over the course of the interaction. Finally, people can share common ground due to physical co-presence – when they inhabit the same physical setting (Clark and Marshall, 1981). Common ground cannot be properly updated without the process called ‘grounding’ (Clark and Brennan, 1991). The term grounding refers to the interactive process by which communicators exchange evidence about what they do or do not understand over the course of a conversation, as they build a mutual pact (Monk, 2003). Grounding is not a static process, it changes during conversations and collaborations. Group members alter their grounding methods according to the purpose for the communication and the medium used to communicate. That means, what two people are trying to accomplish in their communication (purpose), and the techniques available in the medium for accomplishing that purpose, and what it costs to use them. The medium that is best for a purpose depends on the form grounding takes in a medium and whether that serves the participants’ purposes.

Grounding techniques available in one medium may not be available in another, and even when a technique is available, it may cost more in one medium than in the other. Clark and Brennan (1991) identify eight features of two-way personal media and relate how they affect grounding:
copresence: users are near each other, and can point at objects in common ground
visibility: users can see each other; allows gestures, facial expressions
audibility: users can hear each other, and use natural language
contemporality: users can expect to receive a timely reply; interruptions or delays are significant
simultaneity: users can send and receive at the same time; allows interruption, backchannel feedback
sequentiality: user contributions are strictly ordered, and cannot get out of order
reviewability: users can look at the past history of the conversation
revisability: users have the option of editing their contributions before they commit to them

The audio channel has the following features: audibility, cotemporality, simultaneity, and sequentiality. The video channel has the following features: visibility, audibility, cotemporality, simultaneity, sequentiality. This means that only in the video communication modality people have the possibility to share indicative gestures, facial expressions, and body language. This implies, as discussed in section 2.4.1, that common ground will be better supported when using a video channel, as opposed to the use of an audio-only channel in collaboration.

Clark and Brennan (1991) describe how the theory of common ground can be put into practice. As mentioned earlier, in section 2.2.1, many conversations focus on artifacts and their identities. According to Clark and Brennan (1991), mutually establishing references to artifacts plays an important role in building common ground. To successfully establish a reference to an artifact, several techniques are identified. In face-to-face communication one way of referring to artifacts is the ‘indicative gesture’, which highlights the importance of pointing, looking, and touching as a means of grounding references. Establishing ‘referential identity’, which is the mutual belief that addressees have correctly identified a referent, is considered an important issue in CSCW, especially when using media like video, audio and text. For example, when using communication modalities without the visibility feature, e.g. audio, establishing referential identity is not self-evident. The costs for establishing referential identity are higher in mediated communication, in comparison to face-to-face communication. This directly influences the costs for establishing common ground between the collaborators.

2.6.2. Media richness theory
Communication modalities and face-to-face communication are often studied from the media richness theory point-of-view. The media richness theory (hereafter called MRT) proposes that communication modalities have varying capacities for resolving ambiguity, negotiating varying interpretations, and facilitating understanding (Daft and Lengel, 1986). Related to this, the two main assumptions of MRT are: people want to overcome equivocality and uncertainty in organizations, and a variety of media commonly used in organizations work better for certain tasks than others (Daft and Lengel, 1986). Rich media enable people to interpret and reach agreement about difficult and complex issues, while lean media are appropriate for communicating about routine activities.

Using four criteria, Daft and Lengel (1986, p. 560) present a media richness hierarchy, arranged from high to low degrees of richness, to illustrate the capacity of communication modalities to process ambiguous communication in organizations. The criteria for this hierarchy are:
- the availability of instant feedback
the capacity of the medium to transmit multiple cues such as patterns of emphasis and intonation in language, tone of voice and body language
the use of natural language, that is, using language in a natural way, e.g. in conversations
the personal focus of the medium, e.g. a personalized letter as opposed to a general message directed at a group of people

Face-to-face communication is the richest communication medium in the hierarchy followed by telephone, electronic mail, letter, note, memo, special report, and finally, flyer and bulletin (Trevino, Daft, and Lengel, 1990).

However, the theory fails to consider situational elements (e.g. time and place) that might affect behavior, and social factors (e.g. social norms and attitudes) that might shape perceptions of the media. Furthermore, previous research does not provide sufficiently convincing evidence supporting the theory (Suh, 1999). These findings are based on a study by Suh (1999). The author designed a laboratory experiment to examine MRT. A total of 316 undergraduate students took part in the study. The investigation focused on the effect of four communication modalities (face to face, video, audio and text) on task performance and process and outcome satisfaction for both, intellective and negotiation tasks. The negotiating task was divided in two parts: a consonant negotiation task and a non-consonant negation task. A 2 x 4 factorial design incorporating communication modalities and task types was used to answer the research questions. The hypotheses were:

- For the intellective task, decision quality of the text group will be lower than the other groups. The text group requires the most decision-making time, followed by the face-to-face group. Process satisfaction will be higher in the audio and the video groups than in other groups. Outcome satisfaction will be higher in the audio and the video groups than in other groups.
- For the negotiation task, the consonant subjects will be most successful in terms of payoff in the face-to-face group, followed by video, audio and text. The text group requires the most decision-making time, followed by audio, video, and face-to-face groups. Process satisfaction of consonant subjects will be highest in the face-to-face group, followed by video, audio, and text, while process satisfaction of non-consonant subjects will be the reverse. Outcome satisfaction of consonant subjects will be highest in the face-to-face group, followed by video, audio, and text, while outcome satisfaction of non-consonant subject will be the reverse.

Task performance was measured by effectiveness and efficiency. Decision quality – the correctness of the solution for the intellective task and negotiation payoff for the negotiation task – was measured for effectiveness and decision time (the number of minutes required for the pairs to agree on a decision) was employed for efficiency. Task satisfaction was evaluated by a questionnaire that measured satisfaction with the decision process and outcome satisfaction.

In general, the results of the study did not support the hypotheses which were suggested by MRT. Only one hypothesis was supported: “For the intellective task, the text group requires the most decision-making time, followed by the face-to-face group”.

A possible explanation for these results, according to Suh (1999), is that the communication modalities employed in the empirical studies are too similar in terms of richness to differentiate their effects on performance. In this perspective, Suh differentiates between oral and non-oral media. Another explanation, suggested by Suh is that other factors than media richness might strongly affect task performance and satisfaction as well, for example, people may prefer visual contact and openness.
even though these take more time. Also, the subjects may prefer the media which make them feel as if they are using a ‘state of the art’ communication medium.

2.6.3 Previous research on common ground
The theory of common ground proved to be very useful in studying the impact of communication modalities in collaborative work situations. In the following section an overview will be given of previous studies on common ground.

In an experimental study of common ground in text-based communication, McCarthy, Miles, and Monk (1991) found support for their hypothesis that common ground is more difficult to establish when media-features such as co-presence and visibility are not available. In this experiment pairs were asked to use text-based synchronous communication to reach consensus on solving problems with the lay-out of a bank. Twelve pairs of computer science students participated in this study. Common ground was measured by the extent to which partners recalled the ideas they agreed upon, after the collaboration.

In a study by Veinott et al. (1999) the theory of common ground was used to investigate task characteristics and the influence of language used in the collaboration. Common ground was in this study defined as “a shared common language and culture” (p. 303). In this study the participants were divided between native English speakers (NE-speakers) and non-native English speakers (NNE-speakers). To be considered a NNE-speaker in this study, one had to have a native language other than English and have lived in an English speaking country for less than four years. In total thirty-eight pairs of students participated in the study. NE-speakers with NE-speakers and NNE-speakers with NNE-speakers were randomly divided between an audio condition and a video condition. Each pair was to reproduce the path shown on the instructor’s map onto the followers map. In this task participants had to share goals, different information and needed effective communication for good performance. Task performance, subjective ratings of satisfaction, and communication patterns were measured and compared. Task performance was measured by the time to complete a map and the accuracy of the drawing on the follower’s paths compared to the instructor’s path. Verbatim transcripts were made to analyze the communication patterns. A questionnaire was filled out by the participants, after the collaboration, to measure process and medium satisfaction.

The authors expected that video would improve performance and satisfaction as well as changes in the structure of the conversations in the NE-speakers group. Further, they expected that video would change the details of how common ground is established. The authors predicted that pairs in the video condition would have fewer clarification questions and need to check for mutual understanding less often (because the pairs would be able to visually monitor understanding). It was expected that these differences to be greater for the NNE-speakers. The results showed that there was no performance difference in the presence of video for the NE-speaking pairs. However, NNE-speaking pairs performed significantly better with video compared to NNE-speakers in the audio condition. The performance loss in NNE-speaking audio seemed to be associated with less instructing and checking their mutual understanding, as was found in the communication patterns. In the post-experimental questionnaire, NNE-speaking pairs rated the video more useful than NE-speaking pairs. Surprisingly, NE-speaking pairs rated the audio condition more efficient. The less proficient a person rated his or her partner’s English, the more video was valued.
Kraut et al. (2002) studied the use of visual information in shared virtual spaces. Participants in their study shared a visual space consisting of a view of the work area rendered on each of their computer screens. In the role of ‘worker’ and ‘helper’, the participants worked in pairs in completing a jigsaw puzzle. Twelve pairs of undergraduate students participated in this study. Two variables were measured: task performance, measured by time to complete a puzzle, and conversational efficiency, which was analyzed with verbatim coding of the conversations, in which the primary purpose of each utterance was captured in a scheme. Overall, the shared visual space improved performance and conversational efficiency. The authors found broad support for the theory of common ground and its importance for conversations. They also found support for the hypothesis that different communication features change the costs of achieving common ground.

For the sake of this study, the following definition of common ground is used: “For a conversation to be effective the communicators need to understand each other, that means they are noticing and hearing each other, they understand what was said and what was meant”. This definition relates directly to four states which Clark and Brennan identified (1991, p. 224), in which a receiver can be, after the communicator sent a message: 0. the receiver did not notice the message, 1. the receiver did notice the message, but is not in state 2, 2. the receiver correctly heard the message, but is not in state 3, and 3. the receiver understands what was said and the meaning of the message. States 1, 2, and 3 are subject of this study, because these states could be of more interest to collaborators than for example the form of their dialogues.

2.7 Hypotheses
As presented in chapter one, the research question for this study will be:

Compared to the use of an audio communication modality, how is common ground supported by a video communication modality when people are planning a document together with the use of a collaborative writing tool?

First single-author writing and collaborative writing were discussed to give a context for this research question. This led to the discussion of shared artifacts and collaborative writing tools, in which arguments came forward to use the shared writing tool Aspects. Then, the audio and video communication modalities were discussed. Thereafter the activities in the planning stage of a collaborative writing task, and the issues involved with common ground were discussed.

To study grounding, the definition by Clark and Brennan (1991) is:

For a conversation to be effective the communicators need to understand each other, that means they are noticing and hearing each other, they understand what was said and what was meant.

The effectiveness of conversations, in this definition, and the collaboration as a whole, is in this study related to task performance, which will be measured with the quality of the product of the collaboration. The states in which a receiver can be after a communicator sent a message are: 1. the receiver did notice the message, but is not in state 2, 2. the receiver correctly heard the message, but is not in state 3, and 3. the receiver understands what was said and the meaning of the message. ‘Noticing’, ‘hearing’, and ‘understanding what was said’ combined with ‘understanding what was meant’ are used to measure the perceived success of grounding, and will be elaborated in chapter 3, section 3.2.2.
It is expected that in a situation which requires intensive collaboration, e.g. a planning stage in writing, grounding is better supported by a richer medium, in this case, video. As discussed in section 2.5, video has one feature more than audio, i.e. visibility. Visibility supports establishing referential identity and allows indicative gestures, facial expressions, and body language. Therefore, it is predicted that common ground is more difficult to establish when the feature visibility is not available.

To study state 1 “The receiver did notice the message, but is not in state 2” in common ground, it is expected that:

\[ H_1 \quad \text{Participants in the video condition will notice more easily when the partner is about to say something than participants in the audio condition} \]

Participants in the video condition have four sources of visual information with regard to the audio condition. Participants in the video condition have a view of the partner’s head and face, the partner’s upper body and actions (gestures, eye-gaze, and body position). These four sources will influence how well participants can monitor task status, partner’s actions and how well one can identify what the partner is attending to.

To study state 2 “The receiver correctly heard the message, but is not in state 3”, it is expected that:

\[ H_2 \quad \text{Participants in both conditions will hear each other equally well} \]

Participants do not need visual information to hear an utterance or a message and the quality of audio will be the same for both conditions.

To study state 3 “The receiver understands what was said and the meaning of the message”, it is expected that:

\[ H_3 \quad \text{Participants in the video condition will better understand what was meant than participants in the audio condition} \]

Especially in complex collaborative activities, participants in the video condition might benefit from visual information to monitor the partner’s comprehension.

Based on the study by Olson et al. (1995) and Veinott et al. (1999) it is expected that, in the collaboration between non-native English speaking participants:

\[ H_4 \quad \text{Participants in the video condition will be more satisfied with the interaction during the completion of the task than participants in the audio condition} \]
\[ H_5 \quad \text{Participants in the video condition will be more satisfied with the outcome of the task than participants in the audio condition} \]
\[ H_6 \quad \text{Participants in the video condition will produce higher quality outlines than participants in the audio condition} \]

Although previous studied did not find significant differences between the use of audio and video on task performance, it is expected that non-native English participants will perform better using video in a complex writing task.

In the next chapter the design for this study will be described based on the theoretical framework which was described in this chapter.
Chapter 3 Design of the study

In this chapter the design of the study will be elaborated. As described in chapter 2, constructs of common ground, satisfaction with the interaction and outcome and task performance within a planning activity are central to the design of this study. In the first section the choice for the method used in this study will be explained. Then, the criteria for selecting participants will be described. Thereafter, a description will be given of the materials used in this study. Next to that, the procedure of this study will be described.

3.1 Choice for research method

There are several ways to collect data for studying common ground. One of the possibilities is to conduct case studies in which conversations are being analyzed (this includes verbal and non verbal communication). This method has been used in most studies of common ground (see amongst others: Greenspan, Goldberg, Weimer, and Basso, 2000; Kraut et al., 2003; Monk, 2003; Mäkitalo, Salo, Häkkinen, and Järvelä, 2001; Olson et al., 1995; Veinott et al., 1999). The most important advantage of this method is the transcriptions of conversations, which provide very detailed information about grounding during the conversations. On the other hand, this method is very time consuming, especially the transcriptions of every word and utterance the participants make.

The other possibility is to conduct a laboratory experiment without analyzing the complete conversation. In this situation, information about the development of common ground can be collected via questionnaires and observation by the researcher. The data collected is only indirect and to a large extent based on the selfjudgment of the participant, but analyzing this data is less time consuming. Due to limitations of time and resources, the second method will be used in this study.

3.2 Participants

Forty-five participants were recruited for the study. Nine of them took part in the pilot test of procedure and materials. The participants were selected based on criteria described hereafter. All participants were non-native speakers of English. Participants were in the age between 20 and 40 and highly educated, as this supposedly matched the profile of the users of CSCW programs. The participants were drawn from the student body of the Royal Institute for Technology and the Stockholm University. However, some students already started working at the time that they participated. The participants were paired in a way that the participants who collaborated did not know each other beforehand. The pairs had to speak English so the researcher matched people from different countries together. The participants were randomly assigned to the audio or video condition.

3.3 Materials

3.3.1 Task

Inspired by the task designed by Baecker et al. (1993), an introductory task was designed to let the participants write one text (not an outline) about the two of them. This was a short task introducing the participants to the shared writing tool and the communication modality. The introductory task took ten minutes. After completing this task, the participants would begin with the experimental task.
The experimental task required the participants to make an outline about Sweden (see appendix B). This topic was easy to write about, and was neutral in a sense that especially the foreign participants would not be culturally offended. Making the outline, the participants had to discuss content and rhetoric. The participants had to discuss three topics in relation to Sweden. Examples of topics were given in the task description. The participants had to make titles and subtitles, put the topics in the right order and distribute the work (what each of the participants was supposed to do if the outline would be used to write the text). With the help of these task requirements the participants were able to draft an outline.

3.3.2 Questionnaire
The questionnaire used in this study was designed to measure the participants’ perception of grounding and satisfaction, see section 2.7. The questionnaire was filled in after the collaboration. It consisted of three parts; the first part of the questionnaire contained 11 questions, about demographics, computer experience, writing experience and collaborative experience.

The second section in the questionnaire measured the variables attached to the constructs ‘noticing’, ‘hearing’, ‘understanding what was meant’, ‘satisfaction with the interaction’, and ‘satisfaction with the outcome’. The variables were set up by the researcher as follows.

The construct ‘to notice’ consisted of three variables, taken from Bradner and Mark (2001) and Fussell et al. (2003):
- How well a participant was able to monitor task status
- How well a participant was able to monitor the partner’s actions
- How well a participant was able to identify what the partner was attending to

This was measured with questions in the questionnaire about:
- To know when to say something to the partner
- To know when the partner is going to say something
- To notice what the partner is doing, or attending to
- To know if the partner is paying attention

The variable used to measure the construct ‘to hear’ was:
- To hear the words spoken by the partner.

The construct ‘to understand what was meant’ consisted of three variables:
- How well a participant was able to forecast responses based on common ground (taken from Isaacs and Tang, 1993)
- How well a participant was able to understand what was said (taken from Clark and Brennan, 1991)
- How well a participant was able to monitor the partner’s comprehension (taken from Bradner and Mark, 2001; and Fussell et al., 2003)

This was measured with questions in the questionnaire about:
- To understand what the partner is trying to say
- To understand what the partner is saying
- To comprehend the suggestions the partner makes
- To understand the way of reasoning of the partner
- The amount of explanations the partner has to give, after asking for an explanation
- The level of difficulty to repair misunderstandings with the partner
- The extent to which the partner understands the meaning of the words
- To make oneself understood
- To be on the same level of thinking with the partner
The variables used to measure the constructs ‘satisfaction with the interaction’, and ‘satisfaction with the outcome’, were set up by the researcher as follows. ‘Satisfaction with the interaction’ was measured with questions in the questionnaire about:
- How the participants worked together
- The degree to which a person could agree with the partner
- The ‘pleasantness’ of discussing with the partner
- The degree to which consensus could be reached with the partner

‘Satisfaction with the outcome’ was measured with questions in the questionnaire about:
- Satisfaction with the content
- Satisfaction with the form of the document (rhetoric)
- Satisfaction with the division of work

The variable to measure ‘task performance’ was:
- Product quality of the outlines, see section 3.5

The key verbs of the variables, for example ‘understand’, and ‘notice’, were put in an online dictionary to search for synonyms and antonyms (retrieved November 7, 2003, from source: http://www.hyperdictionary.com). These synonyms and antonyms were used to make up different questions about the constructs. Table 1 shows which question numbers are attached to the constructs. The questionnaire used in this study can be found in appendix D.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Questions tied to the construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct 1 “noticing”</td>
<td>12, 13, 14, 15, 16, 17, 21</td>
</tr>
<tr>
<td>Construct 2 “hearing”</td>
<td>18, 19, 20</td>
</tr>
<tr>
<td>Construct 3 “understanding what was meant”</td>
<td>22, 23, 24, 25, 26, 27, 28, 29, 30, 31</td>
</tr>
<tr>
<td>Construct 4 “satisfaction with the interaction”</td>
<td>35, 36, 37, 39, 40</td>
</tr>
<tr>
<td>Construct 5 “satisfaction with the outcome”</td>
<td>32, 33, 34</td>
</tr>
</tbody>
</table>

Questions which were asked but do not belong to any construct, were question 38, 41, 42, 43, 44, 45 and 46. Questions 38 and 41 were about the contribution the participant made and the perceived partners’ contribution to the collaboration. Two questions were asked about the partner, question 42 and 43. Question 44 asked a question about the perceived difficulty of the task and questions 45 and 46 asked questions about the participants’ level of English.

The third part of the questionnaire contained 4 open questions (questions 47 to 50) about the collaboration, the efficiency of the communication and medium and possible technical problems.

Several means were used to avoid validity threats. Questions for the constructs were worded in various ways, for example question 12 and 13: “It was easy to know when I should say something to my partner, when we were about to start the task” and “It was easy to discover when my partner was going to say something, during doing the task”. Questions were also put in negative and positive wordings, for example questions 17 and 26: “I felt ignored by my partner, when I was saying something,
during the completion of the task" and "I could understand my partner’s feelings". Besides, the order of questions was mixed.

To secure the reliability of the questionnaire, multiple questions were used to measure an item, in constructs 1, 2, 4 and 5. For example, question 17 and 21 measure the same item: “I felt ignored by my partner, when I was saying something, during the completion of the task” and “My partner paid attention, when I was saying something, during the completion of the task”. A seven point Likert-scale was used to obtain more precise answers, compared to three or five point Likert-scales.

3.3.3 Set up
The following paragraph describes the set up for the video and the audio condition. In the video condition, participants used a video plus audio modality to communicate, and in the audio condition participants used high quality audio-only. Per room a camera, a TV-monitor, a microphone, and a loud speaker were used, see appendix A. In the audio condition the video monitor would be turned off. The video and audio equipment was provided by the Advanced Media Technology Laboratory.\(^1\)

The video monitor was situated in such a way that the participant only had to glimpse (either to the right or left) to see his/her partner on the monitor, the participant did not need to move his/her head, see figure 1 and 2. The participants would have a view of the head, upper body, hands, and keyboard of their partner. The participants used Mac computers to work with the Aspects program (see chapter 2, section 2.3.1).

\(^1\) The AMT Lab is a research environment for conducting experiments on digital moving images and communicating spaces. The AMT Lab project has been developed by researchers from Media Technology and Graphic Arts within the Department of Numerical Analysis and Computer Science (NADA), at the Royal Institute of Technology (KTH), Stockholm.
Figure 1. Overview of Room 1.
3.4 Procedure

3.4.1 The pilot test
The pilot test was conducted in the same way as the experiments, see section 3.4.2. Nine participants were involved in the pilot test; the researcher participated during one session. During the pilot study the following was changed: Twenty-six questions in the questionnaire were improved in formulation and in the first section of the questionnaire scales were improved; the set-up of the camera and monitor in one of the rooms was adjusted (e.g. repositioning of one of the TV-monitors); and changes were made in the task description. Instead of choosing any topic wanted, the participants could choose only three topics. The number of topics was chosen to make the participants negotiate about at least one of the topics.

3.4.2 The experiments
The participants would be introduced to each other upon arrival. The briefing was held in one of the experiment rooms. After the briefing, one of the participants was brought to another room in the same corridor. The participants would find two papers at their desks, one paper with the task description (see appendix C) and one with a short program instruction.

Figure 2. Overview of Room 2.
After 30 minutes of working the participants were warned that they had ten minutes to finish the document and a second warning was given five minutes later. When the participants had finished the main task, they would fill in the questionnaires. After the participants had finished the questionnaire they received a cinema ticket and were thanked for their time and participation. Most participants appeared to like the experiment, and appeared to find it very interesting.

3.4.3 Data analysis
The answers on the general questions in the first part of the questionnaire were analyzed with frequencies. The mean scores per condition of the second part of the questionnaire, the closed questions, were analyzed with an analysis of variance (hereafter ‘ANOVA’) per question. The scores per question for the specific constructs were added together. The sums for the constructs were analyzed with Independent Samples t-tests. A principal component analysis and Cronbachs alpha per construct were used to analyze the validity and reliability of the questionnaire.

3.5 Task performance
Task performance was measured by the quality of the product of the interaction (see section 3.3.1). The time available during the experiment was fixed and limited, therefore completion time was not measured and taken into account when evaluating the outcome. An explorative method was used to assess the quality of the products of the collaborations, the outlines. In this method reviewers had to sort the outlines according to the quality of the product, from high to low quality. A short interview was held with each reviewer after the sorting task was completed.

In total 6 PhD students reviewed the outlines. Two reviewers took part in the pilot study and 4 reviewers took part in the quality assessment of the outlines. The reviewers were selected on the following criteria. Reviewers had to be experienced in collaborative writing and have experience with the production and reviewing of texts. Therefore, PhD students in the field of communication were asked to participate as a reviewer.

The task of the reviewer was to sort the outlines from high to low quality, see appendix E for the task description. These outlines were the products of the experiments held in this study. While the reviewer was working on the sorting task, the researcher would ask the reviewer questions about the strategy used to complete the sorting task. After completing the sorting task the reviewer was asked three questions:

- Why is this (pointing at the specific outline) the best outline?
- Why is this (pointing at the specific outline) outline in the middle of the pile?
- Why is this (pointing at the specific outline) the weakest outline?

A pilot study was set up to gain more insight in criteria used to sort outlines. Based on these two pilot studies three criteria were used by four reviewers. These criteria were: the general amount of idea units within the 3 topics, the organization/structure of the outline, and the presence of a division of labor.

For each review, the outlines – in total 18 – were divided in three groups of 6 outlines. The first group consisted of high quality outlines, the second of middle quality outlines, and the third of low quality outlines. For each group the number of video and audio outlines was counted. The hypothesis was that more outlines made in the video condition will be in the group ‘high quality’ than the outlines made in the
audio condition. And, more outlines made in the audio condition will be in the ‘low quality’ group than the outlines made in the video condition. The mean ranking and standard deviation were analyzed. A low standard deviation means that the reviewers agreed with each other on the ranking of a particular outline, a high standard deviation means that the reviewers were divided on the ranking of a particular outline.
Chapter 4 Results

In this chapter the findings of this study will be presented. First, an overview will be given of the participants and the characteristics of the participant group will be discussed. Second, the findings for grounding and satisfaction will be discussed. Finally, this chapter will end by describing the findings for task performance.

4.1 Participants

From the total of 45 participants, 36 participants participated in the experiments. All participants were in the age group the researcher was looking for. The youngest participant was 18 and the oldest participant was 36 years old. The mean age was 24. Two-third of the participants were men (n = 23), one-third were women (n = 13). As can be seen in table 2, 14 males and 4 females participated in the audio condition. This skewed division of males and females in the audio condition could cause a validity threat to the results. Thus, results found in the audio condition might be influenced by gender factors.

Table 2
Overview of age and gender in the video and audio condition

<table>
<thead>
<tr>
<th></th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Youngest participant</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Oldest participant</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>n male participants</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>n female participants</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

Most of the participants were students (n = 27), the other 9 participants had a master or bachelor diploma and were either working or doing a PhD program. The level of education was spread evenly over both conditions.

4.1.1 Language skills

Language skills are an important influence on mediated collaboration (Veinott et al. 1999). Therefore, questions were asked about the participants’ competence to express themselves in English. A question was asked about how long a participant lived in an English speaking country as well (question 5, 45, and 46).

Table 3 shows that more people than expected have lived in an English speaking country, at least six participants more than one year. However, the participants who are fluent in English are evenly distributed amongst the video and audio condition. One of the participants had English as his native language. This participant was not excluded, since he was not the only participant who was fluent in English.
### Table 3
**Overview of how long participants lived in an English speaking country**

<table>
<thead>
<tr>
<th>Duration</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3 months</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&lt; 6 months</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&lt; 1 year</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&lt; 3 year</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&gt; 5 year</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>never</td>
<td>13</td>
<td>13</td>
<td>26</td>
</tr>
</tbody>
</table>

Most of the participants judged themselves as competent enough to express themselves in English, see table 4. Only three participants disagreed with this statement, one of them strongly disagreed.

### Table 4
**Overview of perceived English-language skills**

<table>
<thead>
<tr>
<th>Score</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Moderately disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slightly disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slightly agree</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Moderately agree</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

Most of the participants (18 participants) judged their partners as proficient enough to express themselves in English, see table 5. The high level of English language skills predicts a smaller influence of medium on collaboration (Veinott et al., 1999). Therefore, results found in this study might be less significant.

### Table 5
**Overview of perceived partner’s English-language skills**

<table>
<thead>
<tr>
<th>Score</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Moderately disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slightly disagree</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slightly agree</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Moderately agree</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>9</td>
<td>9</td>
<td>18</td>
</tr>
</tbody>
</table>
4.1.2 Experience with computers

The researcher expected that the participants would have, to some extent, experience in working with computers and word-processing programs. A question was asked about the experience with text editors (question 7) and with computers in general (question 8). Participants could score their computer experience on a scale from 1 to 10, 1 indicating no experience at all to 10 expert experience.

As can be seen in table 6 the majority of the participants had much experience with word processing programs. Only a few participants wrote seldom documents on the computer.

Table 6
Overview of experience with writing documents on a computer

<table>
<thead>
<tr>
<th>Experience</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Seldom</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Regularly</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Often</td>
<td>6</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Always</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
</tbody>
</table>

As can be seen in table 7, on a scale from 1 indicating no experience at all to 10 expert experience, 6 participants thought they were not so experienced with computers. It might have been that they have answered this question specifically for Mac computers, they might have more experience on Windows computers.

Table 7
Overview of experience with computers

<table>
<thead>
<tr>
<th>Score</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

Overall, the participants had enough experience with word processing programs and computers to participate in the experiments without major problems.
4.1.3 Experience with collaboration

Questions were also asked about the level of experience with collaborative writing, video collaboration, and audio collaboration (question 9, 10, and 11). As expected most of the participants were not unfamiliar with collaborative writing and planning, see table 8.

Table 8
Overview of experience with planning a document together

<table>
<thead>
<tr>
<th>Experience</th>
<th>Video condition</th>
<th>Audio condition</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Seldom</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Regularly</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Often</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Always</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9 shows that participants scored low on the questions about experience with video and audio collaborations. The meaning of audio and video collaboration in the question was not so well defined. A number of participants was thinking of a broad definition (for example having conversations over the phone with a coworker were seen as ‘audio collaborations’). Whereas other participants had a more narrow understanding of the ‘audio collaboration’ concept (in which they pictured themselves working together as they did in the experiment).

Table 9
Overview of experience with video and audio collaboration

<table>
<thead>
<tr>
<th>Experience</th>
<th>Experience with video collaboration</th>
<th>Experience with audio collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video condition</td>
<td>Audio condition</td>
</tr>
<tr>
<td>Never</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Seldom</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Regularly</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Often</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The researcher expected little experience with video and audio collaboration at forehand. Thus, this was controlled by giving the participants an introductory task to get used to the environment and equipment in the experiment.

Six participants reported the experimental task to be difficult. Overall, participants’ characteristics are equal for both conditions, except for gender.
4.2 Validity and reliability of the questionnaire

The reliability of the constructs was measured with a Cronbach alpha. As can be seen in table 10, the alphas have an appropriate level. Hence, no questions were removed before the analyses.

Table 10
Cronbach alpha per construct

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cronbach alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct 1 “to notice”</td>
<td>.74</td>
</tr>
<tr>
<td>Construct 2 “to hear”</td>
<td>.81</td>
</tr>
<tr>
<td>Construct 3 “to understand what was meant”</td>
<td>.80</td>
</tr>
<tr>
<td>Construct 4 “satisfaction with the outcome”</td>
<td>.69</td>
</tr>
<tr>
<td>Construct 5 “satisfaction with the interaction”</td>
<td>.65</td>
</tr>
</tbody>
</table>

The scores on the closed questions did not have a normal distribution. This finding is a serious threat for the use of t-tests. Therefore t-test were only done on clusters of questions, to approach normal distribution. A principal component analysis showed there is one factor explaining 26% of the variance, this component consisted of questions from several constructs. Further, ten small factors were found, varying from explaining 1% of the variance to explaining 9.5% of the variance. Although the structure of the questions per construct is not supported (see table 1 “Overview of questions tied to the constructs in the questionnaire” in chapter 3, section 3.3.2), constructs will be analyzed with t-tests. An alpha level of .05 was used for all statistical tests.

4.3 Grounding and noticing each other

To report the findings for hypothesis 1 “Participants in the video condition will notice more easily when the partner is about to say something than participants in the audio condition” two questions are central in this section. The first question is: “How easily did the participants notice when the partner was about to say something?” and second “Did the participants in the video condition notice significantly easier when the partner was about to say something, compared to the participants in the audio condition?” To report the findings for noticing, closed questions 12 to 17 and 21 were analyzed. Before any analysis could be done questions 14 and 17 had to be rescaled. These questions were negatively worded in the questionnaire. As can be seen in table 11 on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general between 5 and 6 (slightly agree to moderately agree) on these questions. Thus, to answer the first question, the participants did pay attention to each other and noticed each other moderately well.

Table 11
Overview of means for questions about noticing in grounding

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was easy to know when I should say something to my partner, when we were about to start the task (question 12)</td>
<td>5.72</td>
<td>1.00</td>
<td>36</td>
</tr>
<tr>
<td>It was easy to discover when my partner was going to say something, during doing the task (question 13)</td>
<td>5.47</td>
<td>1.34</td>
<td>36</td>
</tr>
</tbody>
</table>
The question central in the hypothesis is: "Did the participants in the video condition notice more easily when the partner was about to say something, compared to the participants in the audio condition?" To answer this question, the mean scores of questions 12 to 17 and 21 were compared per condition. These means per condition were analyzed with an ANOVA per question and with an Independent Sample t-test for the sum of the questions for this construct. Table 12 shows the differences between the mean scores in the video condition and the audio condition.

### Table 12
Mean scores per condition for noticing in grounding

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th></th>
<th>Audio condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>It was easy to know when I should say something to my partner, when we were about to start the task (question 12)</td>
<td>5.78 1.06 18</td>
<td></td>
<td>5.67 .97 18</td>
<td></td>
</tr>
<tr>
<td>It was easy to discover when my partner was going to say something, during doing the task (question 13)</td>
<td>5.67 1.14 18</td>
<td></td>
<td>5.28 1.53 18</td>
<td></td>
</tr>
<tr>
<td>It was hard to notice what my partner was doing in the document (question 14) Rescaled scores:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the completion of the task, I knew what my partner was attending to (question 15)</td>
<td>5.44 1.15 18</td>
<td></td>
<td>5.39 1.09 18</td>
<td></td>
</tr>
<tr>
<td>My partner paid attention to me in general, while we were carrying out the task (question 16)</td>
<td>6.22 ,81 18</td>
<td></td>
<td>5.72 1.49 18</td>
<td></td>
</tr>
<tr>
<td>I felt ignored by my partner, when I was saying something, during the completion of the task (question 17) Rescaled scores:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My partner paid attention, when I was saying something, during the completion of the task (question 21)</td>
<td>6.67 1.69 18</td>
<td></td>
<td>6.44 .78 18</td>
<td></td>
</tr>
</tbody>
</table>
The participants in the video condition did not score significantly different from the participants in the audio condition on all the questions in this construct. The sum of scores of these questions together did not differ significantly between the two conditions, \( t(36) = 1.120, g = .271 \).

There were no significant differences found between the video and audio condition. Therefore, hypothesis 1 will not be accepted. There were in general no problems starting up the collaboration. Observations of the conversations show that participants in the audio condition sometimes had a little difficulty in starting up the conversation, probably due to the lack of visual information. In the video condition, the participants used visual information from the video monitor to see if the partner was ready to start a conversation. The participants in both conditions did not report difficulties with discovering when the partner was going to say something, during executing the task. One explanation for this finding can be that people tend to listen to their partner to notice when the partner is about to say something, during working with a computer monitor. In other words, the participants in both conditions might tend to focus with their eyes on the computer monitor and listen when the partner is about to say something.

It was easy for the participants to notice what the partner was doing in the document, since it was easy to see the cursor of the partner in the document and the partner’s typed words. Especially, the means on the questions 17 and 21 are more positive than the other questions in this construct, see table 11. This shows that monitoring the task status was well supported by the shared editor. According to findings in open question 50, five participants in the video condition, as well as three participants in the audio condition, reported to have minor problems with the shared editor. The following problems were reported. The arrow keys did not work in the program, it was hard to find out how to change fonts in the program, three pairs had difficulties opening a new shared document, and a spelling checker was missing in the program.

4.4 Grounding and hearing each other

To report the findings for hypothesis 2 “Participants in both conditions will hear each other equally well” two questions are central in this section. The first is: “How well could the participants hear each other?” and the second “Were there differences in hearing each other between the conditions?” To report the findings for hearing, closed questions 18 to 20 were analyzed. As can be seen in table 13 on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general 6 (moderately agree) on these questions. Thus, to answer the first question, the participants could hear each other well.

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>I heard well what my partner was saying to me (question 18)</td>
<td>6.67</td>
<td>.63</td>
<td>36</td>
</tr>
<tr>
<td>My partner heard well what I was saying to him/her (question 19)</td>
<td>6.4</td>
<td>.94</td>
<td>36</td>
</tr>
<tr>
<td>It was easy for me to hear the words my partner spoke, while we were doing the task (question 20)</td>
<td>6.39</td>
<td>1.13</td>
<td>36</td>
</tr>
</tbody>
</table>

The question central in the hypothesis is: “Did the participants in the video condition hear the partner equally well compared to the participants in the audio condition?” To
answer this question, the mean scores of questions 18 to 20 were compared per condition. These means per condition were analyzed with an ANOVA per question and with an Independent Sample t-test for the sum of the scores for this construct. Table 14 shows the differences between the mean scores in the video condition and the audio condition.

### Table 14
Mean scores per condition for hearing in grounding

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I heard well what my partner was saying to me (question 18)</td>
<td>6.61, 0.61, 18</td>
<td>6.72, 0.67, 18</td>
</tr>
<tr>
<td>My partner heard well what I was saying to him/her (question 19)</td>
<td>6.39, 1.04, 18</td>
<td>6.44, 0.86, 18</td>
</tr>
<tr>
<td>It was easy for me to hear the words my partner spoke, while we were doing the task (question 20)</td>
<td>6.39, 1.09, 18</td>
<td>6.39, 1.20, 18</td>
</tr>
</tbody>
</table>

The participants in the video condition did not score significantly different from the participants in the audio condition on question 18 and 19. Notable are the scores on question 20 about hearing what the partner was saying, $F(1, 36) = .000$, $p = 1.000$. The sum of scores on these questions together did not score significantly different for the two conditions, $t(36) = -.210$, $p = .835$.

On question 20 the participants scored similar and overall, the participants could hear each other well. Therefore, hypothesis 2 will be accepted. The audio channel used during the experiments was of a sufficient quality, the participants heard their partners well.

### 4.5 Grounding and understanding each other

To report the findings for hypothesis 3 “Participants in the video condition could better understand what was meant than participants in the audio condition” two questions are central in this section. The first is: “How well could the participants understand each other” and second “Did the participants in the video condition significantly better understand each other than the participants in the audio condition?” To report the findings on understanding, closed questions 22 to 31 were analyzed, as well as open question 49. Before any analysis could be done questions 22, 25, 27, and 29 had to be rescaled. These questions were negatively worded in the questionnaire. As can be seen in table 15 on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general between 5 and 6 (slightly agree to moderately agree) on these questions. Thus, to answer the first question, the participants could understand each other fairly well.
Table 15
Overview of means for questions about understanding in grounding

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>It did take a long time before I understood what my partner was trying</td>
<td>5.78</td>
<td>1.33</td>
<td>36</td>
</tr>
<tr>
<td>to say, when we were starting up the task (question 22) Rescaled score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could comprehend the suggestions my partner made, when we were</td>
<td>6.21</td>
<td>.96</td>
<td>33</td>
</tr>
<tr>
<td>discussing who should do what (question 23)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I understood what my partner was saying, when we were discussing the</td>
<td>6.14</td>
<td>.83</td>
<td>36</td>
</tr>
<tr>
<td>topics for the outline (question 24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was hard to understand the way of reasoning of my partner</td>
<td>5.89</td>
<td>1.37</td>
<td>36</td>
</tr>
<tr>
<td>(question 25) Rescaled score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could understand my partner’s feelings (question 26)</td>
<td>5.06</td>
<td>1.43</td>
<td>36</td>
</tr>
<tr>
<td>My partner needed to explain him/herself more than once, after I</td>
<td>5.19</td>
<td>1.43</td>
<td>36</td>
</tr>
<tr>
<td>asked for an explanation (question 27) Rescaled score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My partner understood the meaning of my words well, when we were</td>
<td>5.67</td>
<td>1.60</td>
<td>36</td>
</tr>
<tr>
<td>discussing the topics for the outline (question 28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had to work hard to make myself understood (question 29) Rescaled</td>
<td>5.47</td>
<td>1.39</td>
<td>36</td>
</tr>
<tr>
<td>score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to be on the same level of thinking with my partner (</td>
<td>4.97</td>
<td>1.89</td>
<td>36</td>
</tr>
<tr>
<td>question 30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to repair misunderstandings with my partner (question 31)</td>
<td>6.08</td>
<td>.91</td>
<td>36</td>
</tr>
</tbody>
</table>

The question central in the hypothesis is: “Did the participants in the video condition understand the partner better compared to the participants in the audio condition?”

To answer this question, the mean scores of questions 22 to 31 were compared per condition. These means per condition were analyzed with an ANOVA per question and with an Independent Sample t-test for the sum of the questions for this construct. Table 16 shows the mean scores in the video condition and the audio condition.

Table 16
Mean scores per condition for understanding in grounding

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>It did take a long time before I understood what my partner was trying</td>
<td>6,22</td>
<td>.94</td>
</tr>
<tr>
<td>to say, when we were starting up the task (question 22) Rescaled score:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could comprehend the suggestions my partner made, when we were</td>
<td>6,35</td>
<td>.70</td>
</tr>
<tr>
<td>discussing who should do what (question 23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I understood what my partner was saying, when we were discussing the</td>
<td>6,39</td>
<td>.61</td>
</tr>
<tr>
<td>topics for the outline (question 24)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The participants in the video condition scored significantly more positive compared to the participants in the audio condition on understanding what the partner was trying to say (question 22), $F(1, 36) = 4.387, p = .044$. The video condition scored also significant more positive than the audio condition on understanding the partner’s feelings (question 26), $F(1, 36) = 14.776, p = .001$, as well as on the frequency of explanations after asking for an explanation (question 27), $F(1, 36) = 4.283, p = .046$.

Three significant differences were found. Therefore, hypothesis 3 is partly supported. The video communication modality is better in supporting the understanding of what the partner is trying to say at the beginning of the task. This is an interesting finding, there was no significant difference found in question 12 “It was easy to know when I should say something to my partner, when we were about to start the task”. The difference between question 12 and 22 is that question 12 asks for knowing when to say what, and question 22 is about understanding what the partner is about to say, when starting up the task. This could indicate it is more difficult to estimate the behavior of the partner in the audio condition. According to the significant difference between the two conditions found in question 26, video seems to support the understanding of the partner’s feelings better. This could indicate that video supports social cues better than the audio communication modality.

Participants in the audio condition scored significantly higher on explaining themselves more than once, compared to the scores of participants in the video condition.

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was hard to understand the way of reasoning of my partner (question 25)</td>
<td>6,28 ,83 18</td>
<td>5,50 1,69 18</td>
</tr>
<tr>
<td>Rescaled score:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could understand my partner’s feelings (question 26)</td>
<td>5,83 1,04 18</td>
<td>4,28 1,36 18</td>
</tr>
<tr>
<td>My partner needed to explain him/herself more than once, after I asked for an explanation (question 27) Rescaled score:</td>
<td>5,72 1,36 18</td>
<td>4,67 1,68 18</td>
</tr>
<tr>
<td>My partner understood the meaning of my words well, when we were discussing the topics for the outline (question 28)</td>
<td>5,61 1,75 18</td>
<td>5,72 ,96 18</td>
</tr>
<tr>
<td>I had to work hard to make myself understood (question 29) Rescaled score:</td>
<td>5,61 1,88 18</td>
<td>5,33 1,94 18</td>
</tr>
<tr>
<td>It was easy to be on the same level of thinking with my partner (question 30)</td>
<td>5,11 1,60 18</td>
<td>4,83 1,65 18</td>
</tr>
<tr>
<td>It was easy to repair misunderstandings with my partner (question 31)</td>
<td>6,28 ,83 18</td>
<td>5,89 ,96 18</td>
</tr>
</tbody>
</table>

The sum of scores on these questions together did not score significantly different between the two conditions, but close to significance: $t(33) = 1.920, p = .064$. 
condition. This finding would suggest that when a participant in the video condition would ask for an explanation from the partner, the partner would give fewer explanations compared to a partner in the audio condition. Needing fewer explanations to repair misunderstandings suggests that the participants can understand each other more easily, thus, sharing a higher level of common ground compared to the participants in the audio condition. This would lead to the conclusion that video compared to audio supports common ground better between collaborators in a writing situation. However, surprisingly there were no significant differences to be found in the majority of questions in this construct. Plausible explanations for this can be that the questions in the questionnaire do not represent the theory well enough or that the theory is not well supported. These explanations will be more elaborated in chapter 5 section 5.1.

4.5.1 Perceived usefulness of the medium
In addition to the closed questions about grounding, open question 49 was asked to gain insight to the perceived usefulness of the medium. Question 49 asked: “Considering the medium you used (“audio only” or “video”), did you find the medium useful in the process of completing the task? Please motivate.”

Twelve participants using video considered the medium to be useful in the process of completing the task:

- 10 video: “I mainly used the audio. I think it is enough to understand but sometime when I am a bit 'lost', I checked with the video”
- 11 video: “Yes. The audio channel was very useful for planning. The video channel was good to see what my partner was doing”
- 14 video: “Yes. When somebody wanted to change the topic, we just had to say it”
- 15 video: “Video was very useful, because you could see what the other person though about what I was saying. He couldn't lie by his voice”
- 22 video: “It was very useful to both hear him and see him; could sense his attitude and feelings…”
- 23 video: “The most important is the audio to be efficient, but the video brings something else. Actually you can see how your partner is working and you lose less time at trying to know what she is doing”
- 30 video: “Yes, I think. Sometimes people are shy to express their dissatisfaction, and one just sees that in their face. For example, my partner did not like the word "openness", because it was strange for her in the context. I noticed it through "video"; otherwise it may not be understood by me”
- 31 video: “Yes it was useful. I think that if you are gonna work with someone and you are not in the same place it's good to see that person. Everything worked well. We could communicate to each other and it's the main thing. I think”
- 34 video: “Yes, I found it useful. It feels that the person is near, and you could ask him and also explain what did you mean. So, it was enough efficient to finish (fulfill) the task”
- 35 video: “Yes. Seeing the person helps a lot both to hear (since lip reading helps) and also to use gestures to make oneself understood”
- 26 video: “Yes, it made it more personal and pleasant being able to see the other person”
- 39 video: “It's useful because it's the same feeling as she was with me. I can see the expressions of the partner, etc. I would see if she was happy or not of our work, if she agree... it was a nice work!”

Five participants in the video condition found that the video was an overkill and that audio-only would have been more useful, surprisingly three participants in the audio condition thought that video might have been an overkill.
Ten participants in the audio condition considered the medium to be useful, six of the participants commented:

- 29 audio: “Yes, since we were placed in two different rooms, it was very useful with the audio medium that helped us. It went faster than it would have done without it. We could hear each others, not only visual”
- 16 audio: “Yes. Audio is the most common way of communicating. We all know how to speak on the phone”
- 21 audio: “It was useful enough to complete the task. The oral communication is the most useful one when doing a task like this”
- 33 audio: “Yes, I could hear him clearly, and there was no delay in the sound. It was almost like working next to each other”
- 40 audio: “Of course. When you hear the voice of another problem it's easier to be convinced about things compared to see them written. It’s also much faster”
- 44 audio: “Yes a lot, especially when both can write at the same time & ask questions to each other. When one got used to that idea work became easy, and after we decided who did what it became efficient as 2 ppl were writing on the task”

However, four participants in the audio condition would have preferred face-to-face communication, as did one participant of the video condition. One participant in the audio condition saw no need for cooperation to complete the task.

The explanations why the medium was seen as useful differed between the two conditions. The participants in the video condition mentioned the usefulness of video to check what the partner was doing, to check mutual understanding, and to check attitudes and feelings during the interaction. The participants in the audio condition mentioned the efficiency (it was faster according to two participants) and the feeling as if they were working next to each other. The participants in the audio condition compared audio with written communication, while the participants in the video condition mentioned the usefulness of visibility in general. This might indicate that visibility supports monitoring task status, the partner’s actions, and the partner’s comprehension better than audio-only does.

### 4.6 Satisfaction with the interaction

To report the findings for hypothesis 4 “Participants in the video condition will be more satisfied with the interaction during the completion of the task than participants in the audio condition” two questions are central in this section. The first is: “How satisfied were the participants with the interaction?” and second “Were the participants in the video condition significantly more satisfied with the interaction than the participants in the audio condition?” To report the findings for satisfaction with the interaction, closed questions 35 to 41 were analyzed, as well as open question 47 and 48. Before any analysis could be done questions 35 and 39 had to be rescaled. These questions were negatively worded in the questionnaire. As can be seen in table 17 on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general a 6 (moderately agree) on these questions. Thus, to answer the first question, the participants were satisfied with the interaction.
Table 17
Overview of means for questions about satisfaction with the interaction

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>We had a hard time working together (question 35) Rescaled score:</td>
<td>6.28</td>
<td>.94</td>
<td>36</td>
</tr>
<tr>
<td>I could agree with my partner most of the time, when we were discussing who should do what (question 36)</td>
<td>6.42</td>
<td>.73</td>
<td>36</td>
</tr>
<tr>
<td>I could, in a pleasant way, discuss with my partner which topics should be included in the outline (question 37)</td>
<td>6.44</td>
<td>.65</td>
<td>36</td>
</tr>
<tr>
<td>It was hard to reach consensus with my partner, in general (question 39) Rescaled score:</td>
<td>6.26</td>
<td>1.01</td>
<td>35</td>
</tr>
<tr>
<td>I was able to discuss with my partner my ideas for the outline (question 40)</td>
<td>6.11</td>
<td>.92</td>
<td>36</td>
</tr>
</tbody>
</table>

In addition, two questions (question 38 and 41) were asked about the contributions of the partners in the collaboration. Table 18 shows on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general 3 (slightly disagree) on these questions. This could indicate that participants contributed equally to the collaboration.

Table 18
Overview of mean scores for questions about the contributions to the collaboration

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>It's mostly my ideas which are used for the outline in the final product (question 38)</td>
<td>3.19</td>
<td>1.47</td>
<td>36</td>
</tr>
<tr>
<td>It's mostly my partner’s ideas which are used for the outline in the final product (question 41)</td>
<td>3.22</td>
<td>1.44</td>
<td>36</td>
</tr>
</tbody>
</table>

The question central in the hypothesis is: “Were the participants in the video condition more satisfied with the interaction than the participants in the audio condition?” To answer this question, the mean scores of questions 35, 36, 37, 39 and 40 were compared per condition. These means per condition were analyzed with an ANOVA per question and with an Independent Sample t-test for the sum of the questions for this construct. Table 19 shows the mean scores in the video condition and the audio condition.
Table 19
Mean scores per condition for satisfaction with the interaction

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>We had a hard time working together (question 35) Rescaled score:</td>
<td>6.50 .79 18</td>
<td>6.06 1.06 18</td>
</tr>
<tr>
<td>I could agree with my partner most of the time, when we were discussing who should do what (question 36)</td>
<td>6.56 .78 18</td>
<td>6.28 .67 18</td>
</tr>
<tr>
<td>I could, in a pleasant way, discuss with my partner which topics should be included in the outline (question 37)</td>
<td>6.50 .62 18</td>
<td>6.39 .70 18</td>
</tr>
<tr>
<td>It was hard to reach consensus with my partner, in general (question 39)</td>
<td>6.29 1.26 18</td>
<td>6.22 .73 17</td>
</tr>
<tr>
<td>Rescaled score:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was able to discuss with my partner my ideas for the outline (question 40)</td>
<td>6.22 .81 18</td>
<td>6.00 1.03 18</td>
</tr>
</tbody>
</table>

The participants in the video condition did not score significantly different from the participants in the audio condition on each question in this construct. The sum of scores on these questions together did not score significantly different between the two conditions, $t(35) = 1.063$, $p = .296$.

There were no significant differences found between the video and audio condition. Therefore, hypothesis 4 will not be accepted. Both groups were quite satisfied with the interaction during the completion of the task. An explanation for this might be that the questions in the questionnaire were not able to measure existing differences between the two conditions. A second explanation might be that there are theoretically no differences between the two conditions. These explanations will be elaborated in chapter 5, section 5.2.1.

4.6.1 Participants’ work style in the interaction
To gain insight in the participants’ work style during the interaction question 47 was asked. Question 47 asked: “Describe how your group has collaborated throughout the writing task.”

Four participants in the audio condition did not fill out this question. In the answers of the participants the following types of work styles could be distinguished:

- A work style in which one of the participants was the leader of the collaboration. This was reported by 2 participants in the audio condition.
- A work style in which one participant was typing and one participant was making suggestions. Generally, the participants would take turns in typing and making suggestions. This was reported by 3 participants in the audio condition and 1 in the video condition.
- A work style in which the participants divided all the work and worked together afterwards. This was reported by 10 participants in the video condition and 3 participants in the audio condition.

Six participants in the video condition and four participants in the audio condition gave a general description of their work style:
26 video: “We collaborated well and understood each others comments + ideas. We had a similar understanding of the task and we were able to write it”

27 video: “It went slowly in the beginning, because we didn't know each other. But then it went well. My partner was kind and nice so it went better and better and was interesting working together”

14 video: “We didn't pay much attention on the writing text. We preferred to talk”

34 video: “We could see and hear each other, so it was not difficult. We could communicate with each other via video and audio systems”

42 video: “We spoke a lot - Before doing/writing we discussed - We used also the video to see expressions etc”

38 video: “silently!”

40 audio: “First was the learning part when we didn't need each other, then was "getting used to" time to feel comfortable with the equipment, then the rest about 50% of the time we could produce something”

41 audio: “Collaboration went fine; we discussed almost everything orally, so we didn't need to make too many corrections to the text”

29 audio: “We collaborated well I think. We had the same humor and the same thought level”

44 audio: “At first it was odd but along the way it began to feel as if the person was sitting beside me looking over my shoulder. The discussion was flowing and it was an enjoyable experience”

One participant in the video condition and two in the audio condition gave the following answers:

15 video: “There were some problems at the start, because there was much talk and not so much talk about the task. We discussed much what we where going to write about but not so much how to write it, in which order and so. And we didn't write a text together, we wrote two different kinds of texts, and it was a mistake”

13 audio: “It was pretty cozy to get to the theme, then I asked B to provide subtopics, then we grouped them in 3 categories. Since he had problems understanding 'outline' I gave him an example by writing a topic, then corrected his English on another topic he wrote alone”

37 audio: “The collaboration was good, I didn't feel my partner was too pushy, to bossy but the audio was not enough for me to express myself”

Although participants worked in several ways together, work style appears not to have influence on the perceived efficiency of the communication and usefulness of the medium.

4.6.2 Perceived efficiency of the communication

In relation with the satisfaction with the interaction, the participants were asked their opinion about the efficiency of the communication. Question 48 asked: “Was the communication with your partner during the task efficient?”

Two participants were not satisfied, one participant reported this in the audio condition and one participant in the video condition:

21 audio: “No. It could be better (video communication or reality)”

38 video: “I do not think so. I wish we spoke more, exchanged more ideas”

There were two participants, in the audio condition, who reported to have problems with speaking English and therefore experiencing inefficient communication with the partner. Two participants reported they had no comparison to answer this question, one in the audio condition and one in the video condition.
Ten participants in the video condition reported efficient communication without further explanation, as did seven participants in the audio condition. Two participants in the video condition reported efficient communication with the following explanations:

- “Especially the audio was efficient to solve minor misunderstandings rapidly”
- “Thanks to the microphone we could comments what we wanted to write, in which order, and this type of communication was very efficient”

Four participants in the audio condition reported efficient communication with the following explanations:

- “It helped a lot in the decision making process & after decisions were made about who was to work on what I attended my own work without thinking too much about what my partner was doing”
- “We understood each other very well and came fast to conclusions and decisions together”
- “It made our task more quickly and easy”
- “Especially thanks to the good quality of the sound”.

Six participants, of which three in the audio condition and three in the video condition, reported the communication during the collaboration was efficient. These participants added a comment to the efficiency of the communication.

- Audio: “Yes, I think so, maybe a bit slow, it took maybe 2 explanations, to understand”
- Audio: “There was no real communication problems between us but I feel we didn't really understand the task”
- Audio: “It was good but he didn't tell me much of what he thought”
- Video: “I think it was satisfactory. “Efficiency” may need more time and practice. It takes time before people reach that point”
- Video: “It was not without problems but the video really helped”
- Video: “Yes! We spoke about what we wanted to write about. But when we were thinking or writing, we couldn't speak. So sometimes nobody didn't speak”

Twenty-seven participants reported that the communication during the collaboration was efficient. Surprisingly, two participants in the video condition report the usefulness of audio. Two participants in the audio condition reported the efficiency of decision making and understanding the partner. However, the comments to the efficiency of the communication show problems as well, especially with understanding the partner and the task.

### 4.7 Satisfaction with the outcome

To report the findings for hypothesis 5 “Participants in the video condition will be more satisfied with the outcome of the task than participants in the audio condition” two questions are central in this section. The first is: “How satisfied were the participants with the outcome?” and second “Were the participants in the video condition more satisfied with the outcome than the participants in the audio condition?” To report the findings for satisfaction with the outcome, closed questions 32 to 34 were analyzed. Before any analysis could be done question 33 had to be rescaled. This question was negatively worded in the questionnaire. As can be seen in table 20 on a scale from 1 (strongly disagree) to 7 (strongly agree), the participants scored in general 5 (slightly agree) on these questions. Thus, to answer the first question, the participants in both conditions were slightly satisfied with the outcome.
Table 20
Overview of means for questions about satisfaction with the outcome

<table>
<thead>
<tr>
<th>Question</th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am satisfied with the content of the document that I and my partner made together (question 32)</td>
<td>5.44</td>
<td>1.59</td>
<td>36</td>
</tr>
<tr>
<td>I am dissatisfied with the form of the document that I and my partner made together (question 33) Rescaled score:</td>
<td>5.00</td>
<td>1.90</td>
<td>36</td>
</tr>
<tr>
<td>I am satisfied with the division of work (question 34)</td>
<td>6.00</td>
<td>1.21</td>
<td>35</td>
</tr>
</tbody>
</table>

The question central in the hypothesis is: “Were the participants in the video condition significantly more satisfied with the outcome than the participants in the audio condition?” To answer this question, the mean scores of questions 32, 33, and 34 were compared per condition. These means per condition were analyzed with an ANOVA per question and with an Independent Sample t-test for the sum of the questions for this construct. Table 21 shows the mean scores in the video condition and the audio condition.

Table 21
Mean scores per condition for satisfaction with the outcome

<table>
<thead>
<tr>
<th>Question</th>
<th>Video condition</th>
<th>Audio condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am satisfied with the content of the document that I and my partner made together (question 32)</td>
<td>5.72 1.36 18</td>
<td>5.17 1.79 18</td>
</tr>
<tr>
<td>I am dissatisfied with the form of the document that I and my partner made together (question 33) Rescaled score:</td>
<td>4.83 2.20 18</td>
<td>5.17 1.58 18</td>
</tr>
<tr>
<td>I am satisfied with the division of work (question 34)</td>
<td>5.89 1.53 17</td>
<td>6.12 .78 18</td>
</tr>
</tbody>
</table>

The participants in the video condition did not score significantly different from the participants in the audio condition on each question in this construct. Surprisingly, table… shows a tendency that the participants in the audio condition were more satisfied with the form of the document and the division of work, compared to the participants in the video condition. The sum of scores from questions 32, 33, 34 did not score significantly different between the two conditions, $t(35) = -.158$, $p = .876$.

There were no significant differences found between the video and audio condition. Therefore, hypothesis 5 will not be accepted. The participants in the video condition were slightly more satisfied with the content of the document, than the participants in the audio condition. However, participants in the audio condition were slightly more satisfied with the form of the document and were slightly more satisfied with the division of work. As for satisfaction with the interaction, two possible explanations for this might be that the questions in the questionnaire were not able to measure existing differences between the two conditions or that there are theoretically no differences between the two conditions. These explanations will be elaborated in chapter 5, section 5.2.2.
4.8 Task performance

To report the findings for hypothesis 6: “Participants in the video condition will produce higher quality outlines than participants in the audio condition”, this section will first describe the extra criteria used by the reviewers in the exploratory sorting method. Second, the problems encountered by the reviewers will be described and finally, the results of the sorting task will be presented.

The three criteria used by the reviewers were: the general amount of idea units within the three topics, the organization/structure of the outline, and the presence of a division of labor. In addition, the following criteria were used by reviewer 1:

- The relative connection between the topics in the outline
- The reason behind the classifications of the topics
- Writing style (logic, coherence, originality)
- The presence of a conclusion to draw the text together

The structure in the text, e.g. the presence of a preview was very important for reviewer 2, while reviewer 3 used in addition the usability of the outline for the reviewer (e.g. could the reviewer use that particular outline to write the full text?). Reviewer 4 added the most extra criteria:

- Writing style
- Perspective of the writer (“I think…”, “We think…”, third person or passive)
- Content
- Target group size (the broader the target group the better it was according to the reviewer)
- Length of the outline
- Characteristics of the writers
- Language use

While working on the sorting task, the reviewers used different work styles. Overall, the reviewers had problems with the diversity of outlines, for example in topics and in the participants’ approaches towards the task, and with the number of outlines.

Overall, the quality of the outlines was not very high. One reviewer in the pilot test put it this way: “Not one is really really good, there are too little good aspects in every outline”. A possible explanation for this might be that the participants were rather focused on the task requirements. As a consequence, most participants did not employ all the activities in planning as discussed in chapter 2, section 2.5. One of the activities neglected by most of the participants was to set goals and sub goals for the writing task. For example, only a minority considered the audience for the document to be written. This aspect of planning was not asked for in the task requirements, but is important to the planning of a document. Also, in a number of outlines the organization and structure for the outline got little attention, as well as in a number of cases the organizational planning, or distributing the work. Two main reasons for this might be the task description which was provided to the participants and the limited time available to the participants. However, this could have influenced the quality of the outlines.

The group of outlines, in total 18, was divided in three groups of 6 outlines. The first group consisted of high quality outlines, the second of middle quality outlines, and the third of low quality outlines. For each group the number of video and audio outlines was counted. The results show that, apart from reviewer 3, there is no difference in quality between the outlines made in the video and audio condition; see table 22. Reviewer 3 used a ‘usability’ guideline, as opposed to the other reviewers. This could indicate that the quality of the outlines is influenced by the usability of an outline. However, the hypothesis that more outlines made in the video condition will
be in the group ‘high quality’ compared to the outlines made in the audio condition is not supported. And, the hypothesis that more outlines made in the audio condition will be in the ‘low quality’ group compared to the outlines made in the video condition is not supported either.

Table 22
Overview of the quality of the outlines in the video and audio condition

<table>
<thead>
<tr>
<th>Quality</th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
<th>Review 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video</td>
<td>Audio</td>
<td>Video</td>
<td>Audio</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Middle</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

The mean ranking and standard deviation were analyzed, see table 23. A low standard deviation (≤ 2.0) means that the reviewers agreed with each other on the ranking of a particular outline, a high standard deviation (≥ 2.0) means that the reviewers were divided on a ranking of a particular outline. As can be seen in table 23, the reviewers agreed on the ranking of outline number 17, 18, 22, 9, 23, 19, 7, 11 and 8. In fact, outline number 8 was put by all reviewers at place 18 in the reviews of the outlines, resulting in a standard deviation of zero. The mean of an outline presents the ranking of the outline on a scale from 1 (highest quality) to 18 (lowest quality).

Table 23
Overview of the mean ranking of the outlines

<table>
<thead>
<tr>
<th>Rank M</th>
<th>SD</th>
<th>Outline number</th>
<th>Audio or video condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.50</td>
<td>.577</td>
<td>17</td>
<td>Audio</td>
</tr>
<tr>
<td>3.25</td>
<td>1.708</td>
<td>18</td>
<td>Video</td>
</tr>
<tr>
<td>4.75</td>
<td>4.924</td>
<td>13</td>
<td>Audio</td>
</tr>
<tr>
<td>5.25</td>
<td>1.500</td>
<td>22</td>
<td>Video</td>
</tr>
<tr>
<td>6.25</td>
<td>1.500</td>
<td>9</td>
<td>Audio</td>
</tr>
<tr>
<td>6.50</td>
<td>3.416</td>
<td>20</td>
<td>Video</td>
</tr>
<tr>
<td>6.75</td>
<td>2.872</td>
<td>16</td>
<td>Video</td>
</tr>
<tr>
<td>9.00</td>
<td>6.218</td>
<td>6</td>
<td>Video</td>
</tr>
<tr>
<td>9.75</td>
<td>2.986</td>
<td>12</td>
<td>Video</td>
</tr>
<tr>
<td>10.00</td>
<td>6.377</td>
<td>21</td>
<td>Audio</td>
</tr>
<tr>
<td>10.25</td>
<td>1.258</td>
<td>23</td>
<td>Audio</td>
</tr>
<tr>
<td>10.75</td>
<td>2.363</td>
<td>15</td>
<td>Audio</td>
</tr>
<tr>
<td>11.50</td>
<td>2.887</td>
<td>10</td>
<td>Video</td>
</tr>
<tr>
<td>11.50</td>
<td>2.646</td>
<td>14</td>
<td>Video</td>
</tr>
<tr>
<td>14.00</td>
<td>.816</td>
<td>19</td>
<td>Audio</td>
</tr>
</tbody>
</table>

47
The participants who made outline number 8 did not pay much attention on the writing text; they preferred to talk (small talk). One of the participants reported: “There were some problems at the start, because there was much talk and not so much talk about the task. We discussed much what we where going to write about but not so much how to write it, in which order and so. And we didn’t write a text together, we wrote two different kinds of texts, and it was a mistake”.

The participants who made outline number 17 worked together very efficient. One of the participants reported that both were engineers and that this background was a help for working together, according to the participant they had the same way of thinking, which meant that they would divide the work as soon as possible and work together when necessary. However, one of them reported also that he missed visual information, which made it difficult to start the collaboration.

Overall, the reviewers agreed on the ranking of 9 out of 18 outlines. There were no significant differences found in the quality of outlines between the outlines made in the video and audio condition. Thus, hypothesis 6 is not supported.

<table>
<thead>
<tr>
<th>M</th>
<th>SD</th>
<th>Outline number</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.75</td>
<td>1.893</td>
<td>7</td>
<td>Audio</td>
</tr>
<tr>
<td>16.25</td>
<td>.957</td>
<td>11</td>
<td>Audio</td>
</tr>
<tr>
<td>18.00</td>
<td>.000</td>
<td>8</td>
<td>Video</td>
</tr>
</tbody>
</table>

Table 23 continued
Chapter 5 Conclusions and discussion

In this chapter the conclusions for this study will be presented and the study will be discussed. First the research question will be answered. Then, the conclusions for satisfaction and task performance will be discussed. Thereafter, the reflections on this study will be presented and to conclude the chapter, suggestions for future research will be given.

5.1 The support of common ground by audio and video

The research question of this study will be answered in this section. The question is: “Compared to the use of an audio communication modality, how is common ground supported by a video communication modality when people are planning a document together with the use of a collaborative writing tool?”

The results of this study show that in the context of planning a document, both audio and video were well able to support the creation and maintenance of common ground. The participants noticed each other, heard each other well, and understood each other. There were very few significant differences between these communication modalities. However, the results show that common ground is supported in different ways by the audio and video communication modalities. In comparison to the video modality, audio supported efficiency well, as was suggested by Sallnäs (2002). Sallnäs suggested that audio might be more efficient compared to video and that video might be a more enjoyable medium. This is supported by the results from especially open question 49, about the usefulness of the medium. The visibility feature in video supported estimating behavior, understanding and feelings, and ‘checking what the partner is doing’ well. Both modalities supported social presence well, as was reported by the participants in the open questions in this study.

The results from observations and the open questions, in particular question 48 “Was the communication with your partner during the task efficient?”, suggest that the participants in the audio condition had more problems in the collaboration, e.g. in starting up the collaboration. However, significant differences were not found in the scores on the closed questions. In chapter 4 two explanations were suggested: the items in the questionnaire might not have represented the constructs well enough, or that differences between the two communication modalities do not exist. Results from the open questions, as well as the significant differences found for understanding clarify that differences in support for common ground do exist between the audio and video modality. A possible explanation is that the items in the questionnaire did not represent the constructs well enough, since this questionnaire was developed for the first time. Two other more probable explanations for the lack of significant differences between the audio and video conditions are the following. As was mentioned in the description of the characteristics of the participant group, language skills might decrease the differences between mean scores for the two conditions. Next to the probable influence of language skills, the use of the shared editor Aspects has influenced the results as well. The shared editor provided, regardless the condition, the participants a visual overview of the task which made it easy to monitor task status and the partner’s actions in the document. These aspects of noticing had a positive influence on the mutual understanding between the participants in the audio condition.

Video and audio appear to support common ground equally well. However, the above arguments lead to the conclusion that in complex situations, in which intensive
collaboration is required, video will support the creation and maintenance of common 
ground better than audio. This conclusion is in concordance with the prediction of 
Isaacs and Tang (1993) that video improves the ability to show understanding and to 
express attitudes and they correspond to the finding that video is especially useful for 
handling interaction-intense activities. Similar to the suggestion of this study that the 
shared editor may have influenced the support of common ground positively, Isaacs 
and Tang (1993) and Olson et al. (1995) mention the importance of integrating video 
with a shared writing tool to collaborate successfully.

5.2 Conclusions for satisfaction
In this section the conclusions will be presented for satisfaction. First, the conclusions 
for satisfaction with the interaction will be discussed, then the conclusions for 
satisfaction with the outcome.

5.2.1 Conclusions and discussion for satisfaction with the interaction
The results found for satisfaction with the interaction indicate that all participants 
were satisfied with the collaboration. There were no significant differences found 
between the audio and video condition for satisfaction with the interaction. This could 
directly lead to the conclusion that using either video or audio would not influence the 
perceived satisfaction with the interaction. Other issues than agreement with the 
partner, the pleasantness of discussing issues with the partner and reaching 
consensus, in collaboration and interaction might influence the satisfaction with the 
interaction. For example, the educational background of collaborators. As was 
mentioned in chapter 4, section 4.8, a participant noted that he shared his 
educational background with his partner and this helped them to work together 
efficiently. Also attitudes and work style might influence the satisfaction with the 
interaction. However, another conclusion might be that the questions in the 
questionnaire were not able to measure differences in satisfaction with the 
interaction, for example as they were found in a study by Olson et al. (1995). Thus, 
the above arguments lead to the conclusion that, in spite of the findings in this study, 
significant differences might exist between users of audio and video.

5.2.2 Conclusions and discussion for satisfaction with the outcome
Unexpectedly, the mean scores of the audio condition on satisfaction with the form of 
the document and the division of work were higher, although not significantly, than 
the mean scores in the video condition. It could be possible that participants who 
worked in the audio condition worked efficiently and therefore were more satisfied 
with the form of the document and the division of labor, rather than the content of the 
outline. Nevertheless, these findings were not significant. Previous research on task 
performance and the use of audio and video did not find significant differences either 
(Kraut et al., 2003; Olson et al., 1995; Sallnäs, 2002; Suh, 1999), apart from the 
study by Veinott et al. (1995). One could conclude that even for complex tasks, in 
which people have to negotiate, differences between the use of audio and video do 
not exist.

Compared to the mean scores on satisfaction with the interaction, the participants in 
both conditions were less satisfied with the outcome. A reason for this can be that the 
participants appeared to like the experiment very well, but were less satisfied with 
what had been produced in the experiment.

The results for satisfaction with the interaction and the outcome are similar to the 
results found in the study by Suh (1999), who investigated the use of media 
according to the media richness theory. The assumption of MRT is that certain media
work better for certain tasks than others, expressed in the satisfaction with the interaction and outcome. This assumption is not supported by this study either. As discussed, other issues in collaboration and interaction might influence the satisfaction.

5.3 Conclusions for task performance

The overall quality of the outlines, which were produced by the participants in this study, was not very high and significant differences between the audio and video condition in outline quality were not found. The results showed that a usability criterion for reviewing the outlines might enhance the differences between outlines made in the audio and video condition.

Previous research did not find significant results for task performance either (Olson et al., 1995 and Kraut et al., 2003). However, Veinott et al. (1999) found significant differences in task performance for non-native speakers of English. The lack of significant differences in the quality of the outlines could be due to the following. The outlines were only reviewed by four reviewers, which is a very small sample-size. The mutual agreement was rather low, the reviewers agreed on the ranking of fifty percent of the outlines. A reason for this might be the problems that the reviewers encountered while sorting the outlines. The criteria for sorting the outlines have also influenced the validity of the method. The criteria which the reviewers had to use might have been problematic, since all reviewers added criteria to sort the outlines. In spite of the pilot test, in which the criteria were set up, the criteria might have been open for individual interpretations and hence, have threatened the validity of the method. Some suggestions for future research on the quality of outlines are:

- Study planning in a field study, in which ‘real work’ is studied, instead of a laboratory experiment.
- Use more reviewers to sort the outlines.
- Limit the number of outlines to sort, e.g. take random samples of the outlines.
- Use the usability-criteria to sort the outlines, ask the reviewers to validate ‘usability’ why a particular outline is perceived as usable.

5.4 Reflections on the method

In this section will be reflected on the method used in this study. First the participants will be discussed. Thereafter, the research method in which a questionnaire was used to measure the support of common ground and satisfaction with the interaction and outcome will be discussed.

The number of participants was sufficient to use statistical analyses, but with regard to generalizing the results of this study the number of participants might have been small. However, this is not a direct threat to generalize the results to a larger population. A greater threat towards generalizing the findings in this study is the way in which the participants were asked to participate in this study. A convenience sample was used; all participants were to some extent familiar to the researcher. The question that rises is how to generalize the results and conclusions? The researcher tried to find participants supposedly matching the population of collaborative writers. However, the experience with collaborative writing and writing in general has been less than expected. This was only observed when the data was analyzed. In addition, the division of gender was skewed in the audio condition, but this was not of influence on the results. Overall, the results for this study apply to non-native speakers of English using audio and video in computer-supported collaborative work. A control group was missing in this study. Therefore, it is not possible to formulate
statements about e.g. native speakers of a language and the actual impact of the
shared editor.

A pilot test was conducted to secure the validity and reliability of the questionnaire. The sample-size used in the pilot test was small. Hence, validity and reliability measures could only be taken after the experiments took place. Therefore, although the principal component analysis did not support the questions tied to the constructs, the data was treated as if the constructs were supported.

Compared to the method in which conversations are transcribed and analyzed, this method is rather practical. The concepts of noticing, hearing and understanding might be more important to collaborators than the form and structure of their conversations. In this study the questionnaire to measure common ground was developed for the first time. Thus, the questionnaire needs to be developed for further use, in spite of the high correlations between the questions and the constructs. The principal component analysis showed that the constructs do not appear to be valid as was proposed by the theoretical framework. This analysis could be further investigated to search for other factors behind the components.

As discussed in the section about task performance, the task represented a planning stage in a writing task not well enough to stimulate the participants to employ most of the important planning activities, such as setting goals for the writing task. Further, the experimental, laboratory setting for this study was unnatural and might have influenced the behavior of the participants. Another influence on the behavior of the participants is that the participants knew the researcher. This makes it difficult to give judgment to the planning stage, as was intended in Chapter 1.

5.5 Suggestions for future research

In this section suggestions for future research will be presented. First, suggestions for replicating this study will be discussed. Second, suggestions for future research on communication modalities and the planning stage of writing will be presented.

5.5.1 Suggestions to replicate this study

To replicate this study the following should be considered. As discussed, the shared editor added value in monitoring the task status and the partner's actions. This might have decreased the differences in noticing and understanding for the audio and video condition. To avoid this pitfall in future, a control group without a shared editor should be added to the study, e.g. the control group should use a text editor which is not shared. Next to the influence of a shared writing tool, language skills could have decreased the differences between the audio and video condition as well. A control group with native speakers could be added to the study. The result would be a 2 x 2 design for the study. As discussed, an important suggestion for future research is the use of more participants, to increase the external validity, especially in a 2 x 2 design.

5.5.2 Future research on the use of communication modalities in planning

As was mentioned at the end of section 5.4 it is hard to give judgment to the planning stage, because it mainly served as a background to study communication modalities and common ground. The planning stage could best be studied in a field study setting, to observe and study the natural behavior of writers. In this setting it would be interesting to focus on the use of communication modalities between novice and expert writers and the creation and support of common ground. Faigley et al. (1985) and Sharples (1999) point out that novice and expert writers use different strategies for planning, composing and reviewing. It could be possible that writing experience
might influence the perception and use of a communication modality. For example, one could hypothesize that expert writers could sufficiently work with audio in a collaborative writing task, whereas inexperienced writers would need video in a writing task to understand each other well enough. In addition, variables as work style and educational background could be studied in a field study on satisfaction with the interaction.

Considering that studying communication modalities is important for the design and development of collaborative writing tools, a recommendation based on this study is that video and audio should be implemented in a collaborative writing tool to support synchronous communication. To conclude this chapter, the theory of common ground proved again to be very useful in studying communication modalities and it is recommended to use this theory when studying communication in all kinds of situations. The theory is useful, because it offers an uncomplicated and practical perspective on communication.
References


Appendixes

Appendix A consists of the audio and video equipment used in this study. Appendix B contains the instructions given to the participants in the experiments. In appendix C the use of the collaborative writing tool Aspects was shortly explained, as the task description. The participants kept this task instruction during the execution of the introductory and experimental task. The questionnaire as was used in this study can be found in appendix D. Appendix E contains the instructions for the reviewers of the outlines, which were given to all four reviewers. Appendixes C to E are in a different character font to represent the documents as they were handed out.
Appendix A: The video and audio equipment

Two video cameras, type JVC KY-19, and two TV-Monitors, type JVC TM-A10E were used, along with two microphones, type Behringer XM8500, and two loudspeakers, type Fostex 6301B were used in this study.
Appendix B: Instructions for the participants

The briefing went as follows: “Well, you are going to work together for the next hour”. At this moment the researcher would point towards the computer monitor and show the participants the program Aspects. The researcher explained they could see each other write in the document. One of the problems of Aspects was explained: the fact that two people cannot have access to the same paragraph at one time. The researcher explained further: “Before you will begin with this task you will be doing an introductory task to get to know the program. It is important that you feel at easy with the program, this is more important than performing the task well. You will have 10 minutes to do this task. After these ten minutes I will give you a warning to proceed with the main task. In the main task you will be planning a document. So you will not write the actual text but merely discuss what topics you want to address in the text, in which order and a global content. For this main assignment you will have 40 minutes. I will mention it when you have ten minutes left and when you have five minutes left. After finishing the task I will bring you a questionnaire to fill in, this will be completely confidential. If you have any questions during doing the task you can ask them at once, because I will be listening with you. I will be sitting next to the haptic lab. The session will be recorded; do you have any problems with that?”
Appendix C: An explanation of the use of Aspects and the task description

To open a new (writing) document:
File \(\rightarrow\) new \(\rightarrow\) document type “writing”; push “new” \(\rightarrow\) new document appears in the window

To save a document:
File \(\rightarrow\) save as \(\rightarrow\) give a name, for example “Yourname+date” \(\rightarrow\) choose desktop as the place to save the document

To open and save a document these are the only things you have to do, don’t worry about settings etc. If you have any questions please ask me.

Task description

First an introductory task

Write together a text in which you introduce yourselves together. For example: name, age, country of origin, why you are in Sweden, education, future plans, family etc. Make one text about the two of you.
Take a few minutes to talk about how to write the text, about the content, and try to finish the text in 10 (ten) minutes.
By doing this task you will get to know the program, to do the next task, practicing is more important than completing this task. Feel free to try out this program, but you are not allowed to use the chat-box, you can only communicate orally.

Our impressions of Sweden

In this task you will make an outline for a text, which means that you are going to make a plan together for writing the actual text. This means I do not ask you to write the whole text, but only a plan for it. The main topic will be about Sweden. Here are some suggestions how to do the task:
You can discuss which topics you want to write about in the text.
For example: the education and health services, the taxes, the public transport, politics, the weather, the Swedish people, buildings, food etc.
You can only pick 3 (three) topics to use in your outline.
You can discuss the order of the topics you will have in the text
You can discuss the title and subtitles in the text
Write short paragraphs to show what the text will be about, for example 4 (four) sentences per paragraph.
You can discuss who is going to write which part of the text, or discuss how to divide the paragraphs amongst each other.

In the end you should hand in a document containing:
- Title and subtitles stating the topics you want to discuss
- The right order of the topics
- The distribution of work (which you would have to do, if the outline would be used to write the text)
Appendix D: The questionnaire

This questionnaire consists of three sections. In the first section general questions will be asked. In the second section questions will be asked on which you can give your opinion by checking a box which matches your opinion as close as possible. The options you can choose are: strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree and strongly agree. The third section consists of open questions, in which you can give your opinion in your own words. There are no wrong or right answers in this questionnaire, your opinion is important. Your answers will be treated confidentially and anonymity will be guaranteed.

Section 1
1. How old are you? …………………
2. Gender: Male / Female
3. Level of education: ……………………………
4. Native language: ……………………………
5. Have you lived abroad in an English speaking country? …
   If the answer is yes: For how long? …………………
6. For how long have you lived in Sweden up to this moment?
7. Do you use the computer to write a document?
   Never
   Seldom
   Regularly
   Often
   Always
8. To what extent do you have user experience with computers?
   On this scale 1 means “no experience at all” and 10 means “very very experienced”
   1 2 3 4 5 6 7 8 9 10
9. I have planned documents together with other persons before
   Never
   Seldom
   Regularly
   Often
10. I have been in an online collaboration using video before
   Never
   Seldom
   Regularly
   Often
11. I have been in an online collaboration using audio before
   
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>Seldom</td>
</tr>
</tbody>
</table>

**Section 2**

12. It was easy to know when I should say something to my partner, when we were about to start the task

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

13. It was easy to discover when my partner was going to say something, during doing the task

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

14. It was hard to notice what my partner was doing in the document

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

15. During the completion of the task, I knew what my partner was attending to

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

16. My partner paid attention to me in general, while we were carrying out the task

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

17. I felt ignored by my partner, when I was saying something, during the completion of the task

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>

18. I heard well what my partner was saying to me

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly</td>
<td>strongly</td>
</tr>
<tr>
<td>disagree</td>
<td>agree</td>
</tr>
</tbody>
</table>
19. My partner heard well what I was saying to him/her

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

20. It was easy for me to hear the words my partner spoke, while we were doing the task

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

21. My partner paid attention, when I was saying something, during the completion of the task

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

22. It did take a long time before I understood what my partner was trying to say, when we were starting up the task

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

23. I could comprehend the suggestions my partner made, when we were discussing who should do what

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

24. I understood what my partner was saying, when we were discussing the topics for the outline

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

25. It was hard to understand the way of reasoning of my partner

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>

26. I could understand my partner’s feelings

<table>
<thead>
<tr>
<th>strongly</th>
<th>disagree</th>
<th>strongly</th>
<th>agree</th>
</tr>
</thead>
</table>
27. My partner needed to explain him/herself more than once, after I asked for an explanation

28. My partner understood the meaning of my words well, when we were discussing the topics for the outline

29. I had to work hard to make myself understood

30. It was easy to be on the same level of thinking with my partner

31. It was easy to repair misunderstandings with my partner

32. I am satisfied with the content of the document that I and my partner made together

33. I am dissatisfied with the form of the document that I and my partner made together

34. I am satisfied with the division of work
35. We had a hard time working together

strongly disagree  strongly agree

36. I could agree with my partner most of the time, when we were discussing who should do what

strongly disagree  strongly agree

37. I could, in a pleasant way, discuss with my partner which topics should be included in the outline

strongly disagree  strongly agree

38. It's mostly my ideas which are used for the outline in the final product

strongly disagree  strongly agree

39. It was hard to reach consensus with my partner, in general

strongly disagree  strongly agree

40. I was able to discuss with my partner my ideas for the outline

strongly disagree  strongly agree

41. It's mostly my partner's ideas which are used for the outline in the final product

strongly disagree  strongly agree

42. My partner was a nice person

strongly disagree  strongly agree
43. My partner was a competent person to work with

<table>
<thead>
<tr>
<th></th>
<th>strongly</th>
<th>strongly</th>
<th>disagree</th>
<th>agree</th>
</tr>
</thead>
</table>

44. I found the task difficult

<table>
<thead>
<tr>
<th></th>
<th>strongly</th>
<th>strongly</th>
<th>disagree</th>
<th>agree</th>
</tr>
</thead>
</table>

45. I was competent enough to express myself in English

<table>
<thead>
<tr>
<th></th>
<th>strongly</th>
<th>strongly</th>
<th>disagree</th>
<th>agree</th>
</tr>
</thead>
</table>

46. My partner was proficient enough to express him/herself in English

<table>
<thead>
<tr>
<th></th>
<th>strongly</th>
<th>strongly</th>
<th>disagree</th>
<th>agree</th>
</tr>
</thead>
</table>

Next page ➔ ➔ ➔
Section 3

47. Describe how your group has collaborated throughout the writing task.

48. Was the communication with your partner during the task efficient?

49. Considering the medium you used ("audio only" or "video"), did you find the medium useful in the process of completing the task? Please motivate.

50. Were there any technical problems?

Thank you for your time!
Appendix E: The instructions to sort the outlines

I will explain what the purpose is of your help. I have 18 outlines from the experiments I held in November 2003. In these experiments I had two conditions: a video condition in which people worked together on an outline using video and an audio condition in which people worked together via audio. I would like you to sort all the outlines in an order from the best outline to the worst outline, regardless the condition they were made in. For that purpose you cannot see which outline was made in which condition.

The criteria you would have to use, to judge the outlines on are:
- the general amount of idea units
- the organization/structure of the outline
- ‘dividing the work’, see task description

After you have finished ordering the outlines the researcher will ask you some questions.

During reading and sorting the outline the researcher will stay with you to answer any questions and to ask you some questions about the sorting process.

Thank you for your help!