Discourse and Hearing Loss: The Effects of a Text Supplement

by

Anita B. Haravon

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2002

UMI Number: 3047225

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

24 April, 2002
Date

LORAINE K. OBLER, Ph.D.
Chair of Examining Committee

2/24/02
Date

ROBERT GOLDFARB, Ph.D.
Executive Officer

JUDITH E. HARKINS, Ph.D.

HARRY LEVITT, Ph.D.

Supervisory Committee

D. JEFFERY HIGGINBOTHAM, Ph.D.

External Examiner

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

DISCOURSE AND HEARING LOSS:
THE EFFECTS OF A TEXT SUPPLEMENT

by

Anita B. Haravon

Advisor: Loraine K. Obler, Ph.D.

Automatic speech recognition (ASR), a technology that converts speech into text in real-time, may prove effective as a method of real-time transcription for people with hearing loss. Although other methods of real-time captioning, such as CART (computer-aided real-time transcription), do exist and are widely used, they require the skills of an additional individual, a highly-trained stenographer or typist. If effective, ASR would free a person with hearing loss from reliance on another individual for communication assistance.

The effectiveness of ASR (Dragon Naturally Speaking, version 4.0) for communication enhancement was evaluated. The Map Task, a collaborative problem-solving task, was used to elicit conversation between six pairs of deaf/hearing communication partners. Deaf participants were all accustomed to producing speech for everyday communication. Participants had conversations under the following conditions: face-to-face, CART-face-to-face, ASR-face-to-face, CART-only and ASR-only. Methods for measuring Communicative Success were drawn from the disciplines of aural rehabilitation and discourse analysis.

Communicative Success was assessed in four ways: Task Outcome, Completion Time, Check Moves and Participant Evaluation. Task Outcome was measured using the Map Task, which was scored for route-reproduction accuracy. Mean map scores across communication conditions were not significantly different. Completion Time was significantly longer in ASR conditions than in the face-to-face-only and CART conditions. Check Moves, which are similar to requests for clarification, were tallied. Significantly more Check Moves were found in the face-to-face-only condition than the CART-face-to-face condition. Participant Evaluation consisted of hearing and deaf participants' ranking the five communication conditions. Rankings were significantly different, with CART-face-to-face ranking first, face-to-face-only, CART-only and ASR-face-to-face tying for second, and ASR-only ranking third. In terms of Transcription Accuracy, ASR was found to be significantly less accurate than CART.

In sum, this study supported the notion that for people with substantial hearing loss, the accurate real-time transcription of conversational speech improved Communicative Success. CART transcription was accurate and was ranked favorably by participants. Accuracy of ASR transcription will need to be improved before it is comparable to CART as a viable as a method of real-time speech-to-text transcription.

DEDICATION

This dissertation is dedicated to the memory of Irving Hochberg, Executive Officer of the Ph.D. Program in Speech and Hearing Sciences, 1974 – 1999. His integrity, dedication and vision for the discipline of Speech and Hearing Sciences have inspired and motivated my own path.

This dissertation is also dedicated to my niece, Serena Haravon Collins, in the hope that she may reach for her dreams and trust in herself and the future.

ACKNOWLEDGEMENTS

This research was supported, in part, by the U.S. Department of Education, National Institutes on Disability and Rehabilitation Research (NIDRR), Grant # H133E80010, to the Lexington School/Center for the Deaf. I want to thank NIDDR for having the vision to support such research.

I want to thank Associate Provost Linda Edwards and the Distinguished Scholar Dissertation Fellowship Committee (2000-2001) for awarding to me the first Irving Hochberg Dissertation Year Fellowship.

So many people at the Lexington School for the Deaf, Gallaudet University and the CUNY Graduate Center helped to see this project to its completion.

From Lexington, I want to thank: Barbara Hantin, Deirdre Knight, Lois O'Neill, King Chung, Mauro Caputi, Eune Hong, Lucianna Parisi, Maria Minolts, Ernie Haeker, Karen Lambert, Michael Ralph, Robert DiMarco, Manny Hernandez, Curtis Inoue, Martha Nemeth, Kathleen Admirand, Adele Agin, George Friebolin, Dee Brodbar, and Chris Ita. I am especially indebted to Julianne Eum, research intern at Lexington. Her assistance with discourse coding, scoring, and inter-rater reliability and her positive attitude were invaluable. A special thank you also goes to Matt Bakke, director of the RERC on Hearing Enhancement at the Lexington Center for the Deaf, who hired me in 1994 as the Dissemination Coordinator. Little did he know that my job at Lexington would help me to find my 'niche' in our field: assistive technology/rehabilitation engineering for people with hearing loss. Matt gave me the freedom to be creative and allowed me to grow as a professional.

From the Communication Studies Department at Gallaudet University, I want to thank: Rosemary Weller, Norman Williams, Paula Tucker, Linda Kozma, Laura Fischer, and the secretarial staff. Their collaboration during the hectic time of data collection and analysis was invaluable.

From the CUNY Graduate Center, Ph.D. Program in Speech and Hearing Sciences, I want to thank faculty, staff , fellow students and alumni, especially: Loretta Walker, Linda Ashour, Gary Chant, Helen Cairns, Arthur Boothroyd, Richard Schwartz, Orna Eran, Doron Milstien, Mira Goral, Peggy Jacobson, Barbara Schmidt, Erika Levy, Pui-san Wong, Hia Datta and Kristen Maul. I'm especially grateful to my current supervisor and colleague, Arlene Neuman, who gave me the latitude and support to get my work done in this hectic last semester before the defense.

I also want to thank the counselors and staff at the CUNY Wellness Center, especially Jennifer Gerber, Fred Stern, Daniel Rothenberg. I'm forever grateful to the wonderfully supportive and courageous members of my Dissertation Completion Support Group and the members of my Self-Help Support Group, which began as a result of one of Fred Stern's Workshops.

Sometimes, a research project can take us on unexpected journeys: Scotland, for example. I want to thank the researchers of the Human Communication Research Centre at the Universities of Edinburgh, Glasgow and Stirling including, Anne Anderson, Alison Newlands, Ellen Bard for sharing their time and expertise with me. I especially thank Gwenyth Doherty-Sneddon for training me in Conversational Games Analysis and answering many questions both in person and via e-mail.

In August 2001, I spent two weeks at the Wellspring House, a writing retreat center in Conway, Massachusetts. It proved to be one of the most productive and reflective two weeks of my life. I want to thank Ann and Preston Browning for creating such a place, full of warmth, creativity and peace.

I have benefited greatly from mentorship by my three committee members, wonderful teachers and scholars, all. First, I would like to thank Harry Levitt, who believed in my abilities, even when I did not. After I withdrew from the Doctoral program in 1993, it was Harry who encouraged (prodded) me to come back and write my dissertation. He had no doubt that I could undertake this endeavor. I am grateful to have had the opportunity to work with him and benefit from his experience and scholarship. He has helped me to understand how to design an experiment, to use statistical tests appropriately and the importance of keeping the bigger picture in mind when analyzing results.

This project could not have happened without Judith E. Harkins' invitation back in 1998. She asked me to participate in state-of-the-art research on automatic speech recognition funded by the NIDRR RERC grant. Judy's generosity with her time and with the resources of her lab at Gallaudet University was incredible. I could not have conducted such technologically sophisticated data collection and analysis without her collaboration. Judy also helped me pin down the theoretical motivation for my research by introducing me to articles on Grounding Theory and the work of researchers at the Human Communication Research Centre in Scotland.

Finally, I would like to thank the chair of my committee, Loraine K. Obler. I first met Loraine in the Fall of 1987, the beginning of my graduate education. The

thoughtfulness, care and time that she devotes to her students continues to amaze me. I have learned so much from her about writing and thinking about research. In recent weeks, several colleagues have asked me for my input on drafts of papers and presentations. Several of them have noted Loraine's influence in my comments and suggestions. I think that is the highest compliment.

I am grateful to my outside examiner, Jeff Higginbotham of SUNY Buffalo, who I met at the RESNA conference in 2001 and who volunteered to read my dissertation. I appreciate his interest in my research and encouragement of my scholarship. His careful reading of my draft and insightful comments were invaluable.

I want to thank my dear friends and family from many parts of my life who encouraged me, cheered me on and helped me emotionally through the challenges of this solitary process. I especially thank Jen Bailes, Liza Bruna, Kim Gladman, Gail Green, Gretchen Hachmeister, and Sue Oren. Sometimes, one is lucky enough to have a friend who goes before you and tests out the waters – so to speak. For me, Linda Jarmulowicz has been such a friend. I watched her struggle and triumph with her own dissertation. Her advice and calming voice were invaluable during my own travails.

Finally, I thank my parents, Aristide and Alda Haravon and my sister, Lea Haravon Collins, for believing in me and supporting me in my journey.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

Introduction

When two people enter into a conversation, how can the success of the interaction be measured? This question can be approached on several levels:

1. Self-perception: How did participants evaluate their interaction?
2. Outcome: Did participants accomplish their goal?
3. Efficiency: How efficiently did participants communicate?

In this dissertation, I will explore these components of communicative success as they apply to automatic speech recognition, an innovative technology that may enhance communication for people with hearing loss.

Communicative success is crucial to people with hearing loss because, often, the simplest exchange of information can lead to extreme frustration and dissatisfaction. Alternative or supplementary techniques, such as amplification, captioning, speech-reading and sign-language interpretation, have been shown to improve communication with varying degrees of success and cost. Since no single communication-enhancement method can be appropriate in all situations, the needs of the individuals and the situation must be taken into account when evaluating the effectiveness of communication-enhancement techniques. For example, decoding speech-reading cues is not a viable option for every individual with hearing loss. For some individuals, supplementing the speech signal with text transcription has been shown to be a reliable and effective method of information transfer.

A new technology, automatic speech recognition (ASR), may prove effective for such individuals. ASR is a computer-based technology that converts speech-to-text in real-time. The benefit to a person with hearing loss is that ASR could provide access to the speech of others via the written word. Although other methods of real-time captioning, such as CART (computer-aided real-time transcription), have been proven effective and are widely used, they require the skills of an additional individual, such as a highly trained stenographer or note-taker. ASR would free a person with hearing loss from reliance on a third individual for communication assistance.

The effectiveness of ASR (Dragon Naturally Speaking, version 4.0) for communication enhancement was evaluated by assessing three levels of Communicative Success: Task Outcome, Communicative Efficiency and Participant Evaluation. Methods for measuring Communicative Success were drawn from the disciplines of aural rehabilitation and discourse analysis.

One aspect of the aural rehabilitation literature has focused on the characteristics of communication breakdowns for people with hearing loss (e.g. Tye-Murray et al, 1994, Caissie et al., 1996). Clinical investigators have designed therapeutic techniques to improve patients' coping skills and to develop strategies for assertive communication. Other research has examined the benefits of speech-reading (understanding speech through visual cues such as lip and facial movement) (e.g. Boothroyd, 1988; Bernstein et al., 1998). Results suggest that although speech-reading can be taught to a certain extent, some people have a skill for speech-reading while others, even with training, simply do not and cannot significantly improve speech-reading skills past a certain point.

The discourse-analysis literature, on the other hand, has investigated the process whereby interlocutors exchange information. Although this literature has not directly addressed the communication difficulties of people with hearing loss, researchers have examined how individuals alter communication strategies to adapt to technological innovation, such as telecommunications and e-mail (e.g. Doherty-Sneddon et al., 1997, Newlands et al., 1998). The theory of *grounding* provides a useful framework to describe these adaptive strategies. *Grounding theory* describes the process of achieving mutual understanding in conversation (Clark & Brennan, 1991). Grounding is achieved when both participants in a conversation are confident that the listener has understood the speaker's intended meaning. This process must be constantly updated throughout any given communicative interaction. Participants update grounding by giving relevant feedback and making appropriate responses.

The notion of communicative efficiency is closely linked to grounding because the cost of achieving grounding in less than optimal situations undermines communicative efficiency. A body of research has compared face-to-face communication with technology-mediated forms of communication, such as telephone (Boyle et al., 1994), video teleconference (Doherty-Sneddon et al., 1997) and live e-mail chat (Newlands et al., 1998). This literature has shown that while individuals have the ability to adapt to new communication technologies and use them successfully (so that communication is grounded), ultimately there is a cost in terms of communicative efficiency. Overall, the findings suggest that face-to-face communication was more efficient than other media because less time and fewer words were needed to complete the same task. In addition, when using communication technology, participants tended to ask questions more frequently to check their own understanding and to resolve miscommunications.

The research methodology cited above was applied to investigate the Communicative Success of ASR and CART for people with hearing loss. At this juncture, the literature review describes earlier research concerning the following topics:

1. Discourse strategies and hearing loss
2. Communication enhancement and communication access for people with hearing loss
3. Grounding theory: implications for technology-mediated communication.

Discourse Strategies and Hearing Loss

Based on clinical observations, Erber (1988) identified communicative difficulties experienced by adults with hearing loss while using speech and hearing as the communication mode. These difficulties include misperceptions, communication breakdowns, unsatisfactory turn-taking and inappropriate sharing of the conversational floor. For example, one partner may do most of the talking or introduce a new topic abruptly because of a misperception of a previous turn. These difficulties often result in superficial and shorter conversations and the frustration of both participants. Erber identified three characteristics of fluent conversation:

- Equal share of the conversational floor between participants (of equal rank).

Share of the conversational floor refers to the amount of time each partner spends actually speaking. It is measured by calculating the average number of words a participant utters per speaking turn, termed MLT (mean length of turn).

- Low occurrence of unrepaired communication breakdowns.

Communication breakdowns occur when one person has not understood the intended meaning of the other.

- Appropriate use of a variety of repair strategies.

Repair Strategies

Repair strategies are ways in which participants attempt to rectify communication breakdowns. Repair strategies are subdivided into specific and nonspecific repairs (Caissie & Rockwell, 1993). A *specific repair* occurs when the listener states specifically what part of a message needs clarification. For example: (1) request for repetition of a specific constituent, (2) request for confirmation of the message, (3) request for an explanation or elaboration and (4) request for confirmation of the topic. *Nonspecific repairs* do not clearly state which part of a message has been misperceived. For example, a listener may simply ask, 'What?' or look quizzically at the speaker.

The type of repair strategy people use has been shown to influence the way they are perceived (Gagné et al., 1991). Listeners tended to perceive specific repair strategies more favorably than nonspecific repair strategies (Tye-Murray et al., 1994). However, participants with hearing loss were more likely to use the nonspecific repair strategy (Caissie & Rockwell, 1993). I suggest that nonspecific repair strategies are perceived less favorably because they are less effective toward establishing *grounding*.

Certain utterances invite a particular type of response from the communication partner. For example, a question will invite an answer. A greeting, like 'Hello,' will invite a greeting in reply. These linked speaking turns are known as *adjacency pairs* (Schegloff and Sacks, 1973). Repair strategies also elicit such paired responses (Tye-Murray et al., 1995). For example, in their study, the *request for repetition of a specific constituent* strategy ('Mary Jane went where?') elicited the specific information requested ('To Australia.'). The *confirmation* repair strategy ('Did you say your mother lives in a shoe?') elicited appropriate feedback ('Yes' or 'No'). The *non-specific* repair strategy ('What? Huh?') elicited a verbatim repetition of the original utterance. Although verbatim repetition was found to be the most common response to the non-specific repair strategy (Caissie & Rockwell, 1993; Jones, Kyle, & Wood, 1987; Lamb, Owens, & Schubert, 1983), restructuring a message rather than simply repeating it was found to be more successful (Gagné & Wyllie, 1989; Schum et al., 1992).

The purpose of an interaction appears to influence choice of repair strategy. When collaborating to complete a task, participants with hearing loss were more likely to use specific repairs (Schum et al., 1992). For open-ended conversation, they tended to use nonspecific repairs (Tye-Murray et al., 1994).

Control of conversational floor

Share of the conversational floor refers to the actual amount of time each partner spends speaking. The way control of the floor is managed is a question of conversational style. Control of conversational floor also depends upon degree of familiarity with the conversation partner. When adult cochlear-implant users conversed with unfamiliar participants, the cochlear-implant users exhibited behaviors associated with a more controlling conversational style (Tye-Murray et al., 1994). These behaviors include longer speaking turns, more interruptions, more fillers (e.g. 'um. uh') and the asking of fewer questions. The study revealed that these behaviors were perceived as uncooperative and elicited unfavorable reactions from communication partners. Analysis of a conversation between a 64-year-old hard-of-hearing woman and her normal hearing friend revealed that the person with the hearing loss uttered more words per turn (Caissie & Rockwell, 1993). The authors concluded that this was indicative of her tendency to control the conversational floor. They also found that a large proportion of communication breakdowns went unrepaired. The authors hypothesized that this was due to a lack of assertiveness in requesting clarification or an unawareness of a misperception. Another study (Caissie et al., 1998) has shown that when conversing with an unfamiliar partner, participants with hearing loss were more likely to take an unequal share of conversational floor. However, the pattern was reversed with familiar partners, whereby participants with hearing loss used more specific requests for clarification, suggesting they were more comfortable admitting lack of understanding in the familiar situation. In addition, during small group interaction, the adults with hearing loss spoke much less than the participants with normal hearing and only made nonspecific requests for clarification.

Theory of Discourse and Hearing Loss

Several factors including familiarity with partner, topic and situation contribute to variations in conversational style among individuals with hearing loss. There has been an attempt to theorize these discourse characteristics in people with hearing loss. For example, Tye-Murray et al. (1995) proposed a model to account for three differing conversational styles of persons with hearing loss: *interactive, passive,* and *aggressive.* The *interactive* style embodies Erber's characteristics of fluent conversation: equal share of the conversational floor, low occurrence of communication breakdowns and appropriate use of repair strategies. The *passive* style describes a person who relinquishes control of the conversational floor and feigns comprehension. The *aggressive* style describes a person who attempts to control the conversational floor and monopolize conversation. Unfortunately, the authors cite no research about the breadth of conversational styles of people with normal hearing. Based on my experience with people with hearing loss, it is clear that indeed some people with hearing loss tend to dominate conversation and make incorrect assumptions about misheard utterances. However, it is unclear if such behavior occurs any more frequently among people with hearing loss, than among the normal hearing population. Such behavior may simply be a reflection of a person's conversational style before the onset of hearing loss. The literature tends to unfairly pathologize people with hearing loss without adequately contrasting them to the rest of the 'normal' population. Thus a model, such as Tye-Murray et al. (1995), that restricts categorization to interactive, passive and aggressive, is limited in its breadth and scope. Moreover, models such as these promote misconceptions regarding the true nature of communication problems among individuals with hearing loss and mask potential differences that arise from some of the same obstacles that characterize breakdowns in normally hearing individuals, albeit to a lesser extent[1]. The model that incorporates grounding theory offers a way to describe communication without value judgement. To consider the communication difficulties of people with hearing loss as a disruption of grounding may allow for a new perspective on this issue.

In summary, people with hearing loss compensate for receiving a degraded signal differently. Some avoid disrupting the flow of conversation by assuming comprehension of a given utterance, rather than asking for clarification. Others attempt to maintain control of a conversation by asking few questions of their conversation partners and mostly talking about themselves. Communication breakdowns such as these may be based on an inability to establish grounding. Participants may not be confident that they have understood the meaning and intent of each other's utterances but the social cost of establishing understanding may be perceived as too high.

Let us now examine the existing means of communication enhancement and communication access for people with hearing loss.

---

[1] Thanks to Peggy Jacobson for her comments on this section.

Communication enhancement and communication access

Perspectives on compensating for hearing loss can be roughly divided into three domains: *culture, rehabilitation*, and *access* (Woodcock and Aguayo, 1999). These areas are by no means mutually exclusive. They are separated here solely for discussion purposes.

Cultural Perspective

The *cultural perspective* considers deafness a trait, not a defect, and promotes the use of *sign language*, a completely visual form of communication. When both participants in a conversation are well-versed in sign language, it is an effective way for deaf or severely hard-of hearing people to communicate. However, it is unlikely and unreasonable to expect that most hearing people will acquire sign language to communicate with deaf people.

Rehabilitation Perspective

The *rehabilitation perspective*, derived from the 'medical model,' diagnoses an individual's hearing and speaking impairments and prescribes treatment. This perspective encompasses both *amplification* and *aural rehabilitation*.

Amplification technology includes hearing aids; assistive listening devices, such as FM, infrared, and induction loop; and cochlear-implants. Aural rehabilitation encompasses speech-reading training, ear training (how to interpret the signal received via residual hearing), and speech therapy (maintenance of speech-production skills) (Woodcock and Aguayo, 1999). It can include assertiveness training, strategies to manage communication breakdowns (e.g. learning to make specific requests for clarification) (Tye-Murray et al., 1994; Erber, 1988) and strategies for maximizing performance with telecommunication (Castle, 1994). The psychological component of aural rehabilitation involves counseling services addressing the mourning for and acceptance of the hearing loss (Woodcock and Aguayo, 1999).

Although these strategies can give some level of grounding, they can never fully replace a non-degraded auditory signal. Without discrediting the contributions of the aural rehabilitation approach for improving communication, the untapped potential of technology should not be overlooked. The following section will describe some developments in the area of technology.

Access Perspective: Focus on speech-to-text

The *access perspective* focuses on changing the external environment to accommodate the needs of an individual with a disability. For people with hearing loss, technologies for access include *assistive listening devices* to amplify the speech signal and *visible media*, such as captioning or sign-language interpreters (Woodcock and Aguayo, 1999). Certain aspects of the *access perspective*, of course, overlap with the *rehabilitation perspective*.

The conversion of the speech signal into text form, or captioning, will be the focus of this discussion and this research project. Captioning can be achieved with varying degrees of speed, accuracy, and completeness. Possible methods of text transcription include:

- Hand/typewriting: speakers transcribe their own speech.

- CART (computer-aided real-time transcription) or CAN (computer-assisted note-taking): a third party does the transcription.
- Automatic speech recognition (ASR): a machine (e.g. a computer) automatically converts speech into text.

Some methods of transcription put the responsibility of the speech-to-text transformation on the speaker (usually the person with normal hearing), for example: handwriting or typing during a face-to-face conversation. More sophisticated methods, such as CART (computer-aided real-time transcription) or CAN (computer-assisted note-taking) require the skills of a professional or semi-professional transcriber. These methods might be more accurately labeled "human-assisted" as well as computer-assisted.

Although transcription of speech to text is fairly common today, much of it is not done in real-time. Most prime time TV programs are off-line captioned. This is done after the program has been recorded, but before it is broadcast. Live programs, like the evening news, however, are captioned in real-time by a CART reporter. Real-time captioning is defined as "the transcription of words that make up spoken language accurately into text momentarily after their utterance" (Stuckless, 1994, p.198). Stuckless asserts that for transcription to be 'real-time', the delay between a spoken utterance (a word or phrase) and its display in print should be less than three seconds.

With most speech-to-text technologies, there is a tradeoff between transcription speed and accuracy. An accurate transcription is free of word-errors such as substitutions, deletions and additions. Typically, as speed of transcription increases, accuracy decreases. Speed and accuracy of transcription also depend upon the talker's clarity and rate of speech and the transcriber's ability. Rate of speech varies from person to person and from one circumstance to another. The rate of spontaneous speech varies from 120 to 215 words per minute (wpm) (Turn, 1974). Stuckless (1994) calculated the lecture rates of 10 college professors. Rates ranged from 112-180 wpm. He calculated his own average handwriting speed to be 32 wpm. Skilled typists can accurately type as fast as 80 wpm. At this time, it is only well-qualified stenotypists who can accurately transcribe speech in real-time. They have the skill to transcribe in excess of 200 wpm with 98% accuracy (Stuckless, 1994). Another issue or limitation is vocabulary size. ASR systems with limited vocabularies may have higher accuracy rates or may be speaker-independent, but only for a limited vocabulary set.

The following sections describe four methods of speech-to-text transcription.

Handwriting

Handwriting is the simplest and earliest method of communicating with deaf and hard-of-hearing people using text. The hearing individual writes or types notes for the deaf person to read. Foster (1992) conducted interviews with 20 supervisors of deaf employees. Eleven of the 20 indicated the use of writing in a supportive or primary role to communicate with their deaf employees. Supervisors indicated that handwriting allowed them to feel confident that deaf employees understood them, i.e. that communication was grounded. Supervisors did complain about the inconvenience and time required as compared to 'normal' face-to-face communication, however. In terms of grounding theory, more effort was required to achieve grounding. Although technologically primitive, handwriting more closely approximates speech than any other system currently available because it is interactive and portable and has no added financial cost. Aside from literacy, it requires no special skills and unlike other text-to-speech systems, require no third party to operate. These advantages often do not outweigh the great effort that must be made by the hearing participant to write every word and, of course, the requirement for both interlocutors to be literate. In addition, handwriting fails in meetings with multiple speakers. A group discussion cannot be effectively transcribed by a participant in the discussion. The transcriber would be hard-pressed to convey more than the topic of conversation. As Woodcock (1992) said, "Access is not provided by making sure the deafened person knows 'they are talking about the budget now.' Access is knowing exactly what is being said about the budget. Access is being able to join in the conversation and contribute (p. 50)."

Computer-Assisted Note-taking (CAN)

CAN (computer-assisted note-taking) is the recording of notes, not verbatim text. It requires a skilled typist/interpreter to take notes and summarize ongoing lectures or meetings. It is also sometimes used in the classroom to transcribe the speech of a teacher. The note-taker must possess excellent typing skills, and the ability to accurately condense and summarize a lecture. The student is able to read the summarizing notes on a computer screen momentarily after they have been spoken. The notes can be saved for later review. The growth and demand for "keyboard skills" in the general population (such course are now offered in schools) has made CAN a more viable option. Some systems, such as C-Print, developed at the National Technical Institute for the Deaf (Stuckless, 1994), increase typing speed by incorporating abbreviations of key words and phrases. An abbreviation for a word is entered and automatically appears in its complete form on the screen. Although this method is good for summarizing a meeting or lecture, it is not "true access" (Woodcock, 1992).

Computer-Aided Real-Time Transcription (CART)

Computer-Aided Real-Time Transcription (CART) or stenographic computer-aided transcription provides real-time verbatim transcription of speech[2]. It enables real-time captioning of live broadcasts, lectures, and meetings. CART relies upon the skills of a stenographer and a stenographic machine linked to a computer. The steno system, in which words and some phrase are transcribed in a single keystroke, allows the stenographer to transcribe in excess of 200 words per minute.

A court reporter uses a stenographic machine in place of a regular keyboard to transcribe speech phonetically in the form of letter codes. Reporters can transcribe faster than 200 wpm and may have personal dictionaries of at least 200,000 words. They can transcribe so quickly because most of the word codes require only a single chord-stroke. The computer then translates the codes into standard orthography.

Both deaf and hard-of-hearing people report that they like having access to CART because the transcription is verbatim and the transcription speed keeps up with the rate of the speaker's productions (Stuckless, 1994). The output screen includes at least three lines of text, so it is possible to scan what was said a few seconds earlier. Students often prefer CART to CAN because the transcription is a verbatim record of a lecture, not a

---

[2] For this discussion, note that I am discussing computer-aided ***real-time*** transcription, in which the stenographer is in the same location as the speakers. This discussion will not address computer-assisted ***remote*** transcription, in which the stenographer listens to a speech or conversation from a remote location and transmits the transcription via telephone lines.

summary (Stuckless, 1994). Disadvantages of CART include cost, equipment needs, and transcription errors. CART is expensive, because the skilled stenographer must be compensated for services rendered. It also requires several pieces of equipment that are somewhat cumbersome. Transcription errors can be avoided if the reporter is given a list of professional jargon terms and proper names to add to the personal dictionary.

Automatic Speech Recognition (ASR)

Automatic speech recognition (ASR) is the transcription of speech by a computer. Unlike CART, ASR does not require another individual to interpret or transcribe the spoken word. Many people with hearing loss have expressed interest in the development of a viable speech recognition device (Harkins & Jensema, 1988). "...Such a device will give people with impaired hearing much greater access to society's prevalent mode of communication -- spoken language. Deaf people in particular will be able to directly and independently 'tune in' to spoken communication in school, in the community, and in the workplace to a degree that is not hitherto possible" (Stuckless, 1994, p.197).

In order for an ASR system to perform effectively and be viable for conversational-style speech, it needs to function much like a human transcriber using a CART system. According to Stuckless (1994), it must transcribe in real-time (at least 200 words per minute), process continuous speech, be highly accurate, recognize a large vocabulary (at least 60,000 words), and understand many different talkers (speaker independence) . It addition, to be useful for everyday communicative interactions, it must be affordable and portable.

Unlike a human transcriber, current ASR technology cannot differentiate between 'communicative' speech and 'non-communicative' speech, such as hesitations, dysfluencies, and extraneous sounds[3]. In addition, it is *speaker-dependent*, meaning that it must first become familiar with a particular speaker's voice before accurate transcription is possible. Transcription is reported to be 95% accurate for dictation-style speech under optimum conditions and has been found to be much lower (59 –89%) for unrehearsed speech, primarily due to dysfluencies and variability in pronunciation (Stuckless, 2000).

Commercially available ASR systems, such as Dragon Naturally Speaking and IBM Via Voice, are able to transcribe continuous speech and recognize over 100,000 words,[4] and are speaker-dependent. According to Stuckless, "Today, some ASR systems are quite capable of processing continuous speech. Other systems can recognize very large vocabularies. Yet others demonstrate a high level of speaker independence. The challenge now is to incorporate all three of these capabilities into a single system" (Stuckless, 1994, p.222).

Knowing the drawbacks of current technology, this study will investigate how well ASR works for conversation-style speech, comparing it to the acceptable performance level of CART. Recent technological advances may give back to the person with hearing loss some aspects of the speech signal that they are missing and thus enable the individual to communicate more effectively. In this project, I propose to evaluate the ability of new technology to provide some aspects of the speech signal and communication cues that the person with hearing loss is not receiving.

---

[3] Software developers are working on this problem (Baig, 2000).
[4] The Oxford English Dictionary contains approximately 500,000 words.

Table 1 summarizes the attributes of each text-supplement method discussed.

	Text Supplements			
Attributes	Hand/typewriting (by speaker)	CAN	CART	ASR
Speed	Slow	Slow	Fast	Fast
Accuracy	High	High	High	Variable
Completeness	No	No	Yes	Yes
Ability to ignore speech production errors	Yes	Yes	Yes	No
Cost per use	None	Lower than CART	High	Low (after equipment purchase)
3rd Party needed to operate	No	Yes	Yes	No
Agency/Privacy	Yes	No	No	Yes

Table 1: Attributes of speech-to-text supplements

Now, let us examine how grounding theory can help us to evaluate different communication media and the effect of assistive technology on communication for people with hearing loss.

Grounding theory

Introduction

Communication is grounded when both participants in a conversation believe that the listener has understood the speaker's intended meaning (Clark & Wilkes-Gibbs, 1986). Grounding is achieved through "the ongoing collaboration between participants from the initiation of each utterance to its mutual acceptance" (Clark & Brennan, 1989). For example, speakers generally require confirmation that a message was, in fact, received and understood. Listeners, therefore, give feedback about the status of comprehension. Furthermore, it is the responsibility of the listener to request clarification or elaboration of an unclear message. When mutual responsibility for communication does not occur, participants are likely to feel dissatisfied with the interaction. For example, if A is not paying attention (or not cooperating), B may feel that too great an effort is required to communicate. Conversely, if B is not contributing enough information, A may feel the need to make excessive effort to comprehend the information. Note that feedback may also be erroneous.

Let us now examine how participants give evidence that communication is grounded.

Evidence of Grounding

B communicates that she has understood A's contribution is through *positive evidence*, such as an *acknowledgement*, a verbal or nonverbal signal that the listener has understood and the speaker can safely continue. Non-verbal acknowledgements can be gestural signals such as a head nod, a smile, or simply continued attention/eye gaze toward the speaker. Verbal acknowledgements can take the form of *back-channel* responses such as *uh-huh* and *yeah* or *assessments* such as *gosh*, *really*, and *wow*. Another form of positive evidence is the *relevant next turn*, such as an answer to a question, or a response to a request.

If B has not understood A, then she gives *negative evidence*, such as a puzzled look, an irrelevant next turn, or a request for clarification. In addition, interruptions, breakdowns, and repairs are an integral part of the grounding process because they provide information about a participant's knowledge state.

Clark hypothesized that techniques for establishing grounding will vary with the communication medium. For example, face-to-face communication is rich in gestures and facial expression. These visual cues help in establishing grounding. During telephone communication participants learn to compensate for the lack of these visual cues. Clark identified eight grounding constraints or features to differentiate between communication media. In the section below we discuss them in detail.

Grounding Constraints

In terms of communication medium, Clark & Brennan (1991) identified eight *constraints* on grounding. I prefer to think of them as grounding *features*. Consider Clark & Brennan's constraints and how they apply to a variety of communication media. Table 2 provides a summary of the grounding features.

Grounding Features	Communication Media						
	Face-to-face	**Video Teleconf.**	**Telephone**	**Voicemail**	**Walkie-talkie**	**TTY**	**E-mail/ Fax/Letter**
Copresence	YES	NO	NO	NO	NO	NO	NO
Visibility	YES	YES	NO	NO	NO	NO	NO
Audibility	YES	YES	YES	YES	YES	NO	NO
Cotemporality	YES	YES	YES	NO	NO	YES	NO
Simultaneity	YES	YES	YES	YES	NO	NO	YES
Sequentiality	YES	YES	YES	NO	YES	YES	NO
Revisability	NO	NO	NO	YES	NO	NO	YES
Reviewability	NO	NO	NO	YES	NO	YES	YES

(based on Clark & Brennan, 1991)

Table 2: Grounding features for different communication media

1. **Copresence**: A & B share the same physical environment (face-to-face communication).
2. **Visibility**: A & B are visible to each other. This is a feature of face-to-face communication and video teleconferencing (video telephone). During a video teleconference, participants can see and hear each other, although they are not copresent.
3. **Audibility**: A & B can hear each other. This is a feature of face-to-face, telephone, teleconference, and voice messaging. Voice messaging technology allows A to record an audio message on a home answering machine or a computer-based voice mail system for B to listen to at another time.

4. **Cotemporality:** B receives a message about the same time that A produces it. During face-to-face communication, there is virtually no time delay between production and reception of an utterance. For the telephone and the video teleconference, cotemporality depends upon the quality of the transmission. Some technologies create a delay between production and reception of an utterance. Voice messaging and text-based media, such as e-mail and 'snail-mail,' are *not* cotemporal because B receives a message sometime after A has produced it.
5. **Simultaneity**: A & B can send and receive at the same time. During face-to-face communication, while A is talking, B can give feedback nonverbally, with a nod, or a puzzled look. On the telephone, B can use audible back channels like "uh huh" or "yes" without interrupting the flow of A's message. For voice messaging and text-based media, although both participants can produce messages simultaneously, the messages are usually not cotemporal. In fact, these media are designed for messages to be received sometime after they have been sent.
6. **Sequentiality**: A & B's turns are always in sequence. This is a usually feature of both face-to-face and telephone communication because communication is cotemporal. Participants observe each other sending and receiving utterances in real-time. Problems with sequentiality often result in misunderstandings and communication breakdowns. For example, when A relates the plot of a movie to B, A needs feedback from B to maintain grounding. If A does not wait for this feedback, then, in a sense, utterances will not be sequential because eventually B will need to ask for clarification or repetition of previous utterances.

7. **Revisability**: A can revise a message before sending it to B. (e-mail. letters, some voice mail).
8. **Reviewability**: B can review A's message before responding (e-mail, letter, and voice mail). Written communication media, such as letters, faxes, and e-mail, are *revisable and reviewable*. Unlike face-to-face and telephone communication, participants can revise and edit messages before sending them. In addition, participants can review and even annotate messages before replying. Although messages can be transmitted simultaneously, they lack *sequentiality*. In other words, participants can send messages to each other at the same time, but they may get out of sequence and mixed up. Interestingly, voice messaging is similar to written media in that the receiver of a message can review it before replying and messages can get out of sequence. In the newer, digital forms of voice messaging (voice mail), a person can review and revise a message before sending it (or choose not to send it).

The next section will review research methodology and findings comparing the grounding features of face-to-face communication with several communication technologies. The studies will show how communicative efficiency and grounding are closely linked. Then, the effects of hearing loss on grounding will be considered. Finally, the possible benefits of communication-enhancement technology (ASR and CART, in particular) will be explored.

Grounding Features for People with Normal Hearing

Face-to-face communication

Face-to-face communication is the earliest and most basic form of communication. It requires no technology and can be considered the benchmark for the performance evaluation of technologically innovative communication media. Face-to-face communication has available the greatest number of grounding features (see Table 2). As the next section will illustrate, the presence of these particular grounding features make face-to-face communication the most efficient. During face-to-face communication, participants are in each other's presence (*copresence*). Participants with normal hearing can see (*visibility*), and hear (*audibility*) each other. Their utterances are communicated with no time delay (*cotemporality*). They can send and receive messages in the same instant (*simultaneity*). See Table 2. For example, A can talk while B nods, smiles or looks puzzled in response. It is sometimes even appropriate for A and B to speak at the same time, as in the use of back-channels (verbal indicators of comprehension, such as "yeah" and "uh huh"). When A tells a story, B may utter phrases such as "Oh, my goodness" or "That's terrible." These phrases can overlap with A's speech without disrupting the flow of A's story. In fact, these back channels actually serve as grounding markers. This is one of the ways A knows that B understands her speech[5].

---

[5] Note that the amount of acceptable overlap is culturally defined. See Tannen (1984).

Audio-only communication

The most common form of audio-only communication is the telephone. During telephone communication, "normal" participants can hear each other (*audibility*) without a time delay (*cotemporality*) and they can send and receive contributions *simultaneously* (e.g. back channels). The difference between audio-only and face-to-face communication is the lack of *copresence* and *visibility*. Studies have shown that the lack of these features does, in fact, increase the collaborative effort necessary to establish grounding (Boyle et al., 1994, Doherty-Sneddon et al., 1997). For example, in Boyle et al., 1994, sixty-four participants completed the *Map Task*[6] in face-to-face and audio-only settings. Although, in both conditions, the task was completed in the same amount of time and with equal success in terms of task completion, discourse analysis suggested that more *effort* was required in the audio-only condition to achieve grounded communication. Participants uttered significantly more words and took more turns per dialogue in the audio-only condition than in the face-to-face condition. They also produced significantly more verbal back channels, formal handovers of speaking turns, and interruptions. The authors argue that these differences result from participants' attempts to compensate for the absence of *visibility* and *copresence*. Visual cues communicate a great deal about the state of groundedness of a conversation. During face-to-face communication, participants can monitor their partner's facial expressions and body language. In the audio-only condition, participants, apparently, compensated for the absence of visual

[6] The *Map Task* (Brown et al., 1984) is a collaborative problem-solving task with a clear goal and a well-defined measure of Communicative Success. Each participant in a pair is given a simple map of the same location, except that a few landmarks differ slightly in what they are (e.g. tree or house) or where they are located relative to other landmarks (e.g. below the stream or above the mountain). One participant, the information giver (IG) receives a map with both landmarks and a route drawn on it. The other participant, the information follower (IF) receives a map with landmarks only. The participants work together and reproduce the information giver's route on the information follower's map.

cues by producing more auditory feedback in the form of back-channel expressions. Facial expression and body language also provide cues as to when a speaker is ready to relinquish the conversational floor. In the absence of these visual cues, participants produced more explicit handovers of conversational turns, for example, "Well, that's about all I have to say about 'X.' What do you think?" More interruptions occurred, presumably because more communication breakdowns occurred, suggesting that attempts to compensate were not always successful. The compensatory strategies described above for audio-only communication resulted in a larger number of words produced per dialogue, i.e. increased effort and reduced efficiency.

Examples from "real life" can further illustrate speakers' ability to compensate for the constraints of different communication media. Consider the first few times that children use the telephone. They often nod and smile in response to questions from 'Grandma', not fully realizing that Grandma can hear, but cannot see them (personal observation). Communication via a walkie-talkie or CB-radio is also audio-only. Unlike the telephone, participants cannot make contributions simultaneously because only one participant's audio signal is transmitted at a time. Users adapted to this constraint by developing ritualized language, such as "over" or "over and out" to formally regulate turn-taking.

For deaf people, the teletypewriter (TTY) is the functional equivalent of the CB-radio. It is used instead of the telephone and requires access to the telephone system. The TTY is a text-only communication device that sends messages through telephone lines. Like the walkie-talkie or CB-radio, described earlier, the TTY lacks the *simultaneity* feature. Only one person's text message can be transmitted at one time.

Therefore, TTY users rely upon similar conventions. They type "GA", meaning, "go ahead" to handover a turn and "SK" (stop keying) to indicate that the conversation is over.

The quality of the audio signal also contributes to the ease of communication. Users will only stand for a certain amount of distortion or instability of a signal before deciding that the cost of using an unreliable communication medium is too high. When a new technology is introduced, the audio quality is often poor and cost is high. For example, the early telephones had poor signal quality (Rutter, 1987) and were mainly seen as a tool to transfer information. As the sound quality improved and the cost went down, telephones became increasingly popular for chatting and maintaining relationships. Similarly, early TTY's were expensive and were used mostly for commerce until they became more affordable.

Video-mediated communication

Similar to cellular telephone technology, video-mediated communication (VMC), e.g. video teleconferencing, is a communication medium still in development. VMC systems are used for one-to-one conversations as well as group discussions and meetings. VMC enables participants to see (*visibility*) and hear each other (*audibility*) thus providing a *virtual copresence*. Some methods of video transmission introduce a slight time delay so that utterances are not perfectly *cotemporal*, i.e. they are not produced and received in the same instant. Other transmission methods do not allow for *simultaneity* and constrain participants to offer contributions one at a time. This is similar to the audio-only walkie-talkie situation and some overseas satellite-based telephone connections. In some cases, the video camera angle does not allow for mutual eye gaze so participants do not feel that they are looking at each other (O'Conaill & Whittaker, 1997). This reduces any illusion of *copresence*. The ability to simulate copresence is improving with technological advances. The current trend in video-conferencing is toward 'seamless' communication, i.e. a more realistic approximation to copresence.

Research has suggested that the lack of *copresence* increases the collaborative effort necessary to establish grounding (Doherty-Sneddon et al., 1997). In experimental settings, face-to-face communication has been demonstrated to be significantly different from both audio-only and VMC on several measures. Face-to-face required fewer words, turns, and interruptions to achieve the same level of performance as the other conditions. No significant differences were found between audio-only and VMC. VMC tended to be more formal, with fewer interruptions and more words per turn than face-to-face communication (Anderson et al., 1997; Doherty-Sneddon, et al., 1997; O'Conaill & Whittaker, 1997). This effect was especially strong when the technology introduced a noticeable time lag between presentation and reception of audio and video signals (O'Conaill et al., 1993).

Computer-mediated (text-based) communication

Computer-mediated communication (CMC) is text-based communication in a computer "chat" environment, i.e. participants send each other typed messages via computer. Participants are not *copresent* and cannot *see* or *hear* each other. Messages are *cotemporal*, i.e. they are sent instantaneously; and can be *simultaneous*, i.e. communication partners can send and receive messages at the same time. Messages can be edited, so therefore communication is *revisable* and *reviewable*. Note that messages may be received out of the intended *sequence*. One might expect grounding to be more effortful than face-to-face communication because auditory and visual feedback are unavailable.

Newlands et al. (1996) examined how users adapted to this novel, technology-mediated environment. Participants completed the *Map Task*, the task used by Boyle et al. (1994) to compare face-to-face and audio-only communication. On the first trial, participants performed poorly. Some pairs were unable to complete the task. By the third trial, however, performance had improved to equal that of the participants in the face-to-face and audio-only trials of Boyle et al. Interestingly, the text-based dialogues were significantly shorter, in terms of linguistic output (word count), but considerably longer in terms of duration. CMC trials took, on average, one hour compared to the ten minute average for the spoken versions. Newlands et al. speculated that that the CMC task required more time was because the act of typing is more time consuming and more effortful than speaking. Fewer words were produced in the CMC environment because the cost and effort to produce a typed word is significantly higher than the cost of producing a spoken word. In a sense, each word produced had a higher effort value in this text-based medium.

Unlike face-to-face and audio-only environments, listener feedback was rarely produced or elicited in the CMC environment. One likely explanation is that eliciting and producing feedback are more effortful and time consuming in text-based communication. Apparently the benefit, securing up-to-the-minute grounding, was not deemed to be worth the amount of effort. Surprisingly, Newlands et al. reported that this reduction in feedback did not adversely effect task outcome.

An analysis of questions posed revealed that participants tended to ask *specific* questions in the CMC environment and more *open-ended* questions during face-to-face communication. For example, the information giver would more likely ask, "Do you have a 'White Mountain' in the upper right-hand corner?" rather than, "What do you have in the upper right-hand corner?" Open-ended questions were, presumably, an inefficient way of interacting in CMC.

Summary of Previous Research

In the above studies, compensatory strategies enabled participants to achieve equivalent levels of Communicative Success on the Map Task in all three communication media (audio-only, VMC and CMC) as compared to face-to-face. The technology varied in terms of communicative efficiency and collaborative effort. Video-mediated communication, at its best, shares all grounding features with face-to-face communication except true *copresence*. More importantly, more collaborative effort was required for both audio-only and VMC, than for face-to-face, supporting the notion that a communication medium with the most communicative features (face-to-face) requires the least collaborative effort.

ASR and other forms of captioning are similar to VMC in that issues of time lag between the production and transcription of speech are relevant. Therefore, one might expect to find similar compensatory strategies in ASR-assisted communication, such as formal handovers of turns and fewer interruptions.

Computer-mediated communication is a text-based medium that lacks the grounding features: *copresence, audibility* and *visibility*. Certain grounding strategies, such as back channels, were not as productive in this medium. In particular, in the CMC environment, results suggested that fewer words were necessary to achieve Communicative Success and specific questions were more productive than open-ended questions. ASR and CART are similar to CMC in that they are text-based media. It is likely that similar compensatory strategies (i.e. reductions in listener feedback and increase in specific questions) will be needed to achieve grounded communication.

Grounding Constraints/Features for People with Hearing Loss

Hearing loss can result in communication that is effortful, inefficient and even unsuccessful, due to an impaired auditory channel. To give us further insight into the nature of these communication difficulties, let us examine how the grounding features apply to an aural face-to-face conversation between a deaf person (severe-to-profound hearing loss) and a person with normal hearing. In this situation, the person with normal hearing has access to all the features of face-to-face communication previously discussed: *copresence, visibility, audibility, cotemporality, simultaneity,* and *sequentiality* (Table 3). The deaf person has equivalent access to all the features mentioned above, except *audibility*. This feature, which is absent or weak depending on degree of hearing loss, will affect the ease with which participants can achieve grounding. The deaf person is likely to compensate by speech-reading, thus relying more heavily on the *visibility* feature. Lack of audibility also affects the hearing individual in such pairs. Hearing participants must alter their style of communication by speaking slowly, clearly, and facing the person with hearing loss when speaking. In other words, for the conversation to have any chance at success, both participants must expend more effort. In addition, more communication breakdowns and repairs will result in less efficient communication and possibly participant frustration. The addition of a text-supplement, such as ASR or CART, may provide compensation for the difficulties of hearing loss.

	Face-to-face		ASR/CART-Face-to-face		ASR/CART-only	
	Deaf	**Normal Hearing**	**Deaf**	**Normal Hearing**	**Deaf**	**Normal Hearing**
Copresence	YES	YES	YES	YES	YES	YES
Visibility	YES	YES	YES	YES	NO	NO
Audibility	Partial/ None	YES	Partial/ None	YES	Partial/ None	YES
Cotemporality	YES	YES	Delay	YES	Delay	YES
Simultaneity	YES	YES	YES	YES	NO	· YES
Sequentiality	YES	YES	YES	YES	YES	YES
Revisability	NO	NO	NO	YES	NO	YES
Reviewability	NO	NO	YES	NO	YES	NO

Table 3: Grounding features for speech-to-text supplements

Let us now examine how the addition of a text supplement, such as ASR or CART, could affect Communicative Success.

Face-to-face communication: the effect of a text-supplement

In this study, ASR and CART provided a real-time or near real-time text version of spoken utterances to the participant with hearing loss. ASR and CART were contrasted because CART is known to have excellent transcription accuracy, while ASR transcription-accuracy under these experimental conditions was unknown. The normal hearing participant received auditory and visual information. The person with hearing loss, on the other hand, was required to divide attention between two sources of visual information: the talker's face and the screen displaying text output. Consequently, the grounding features were changed in several ways. The participant with hearing loss saw the hearing person's face and could read all the utterances. The participant could use the text output to check comprehension, i.e. access to the feature *reviewability* (Table 3). There was, however, a slight delay between the speech production and the text transmission that affected *cotemporality* and *sequentiality*. How this impacts upon Communicative Success is unclear, at present. In addition, the effect of dividing attention between the hearing person's face and the text output is also unknown. The success of the text-supplement depended, in large part, on the accuracy of the transcription. In this study, CART was predicted to be more successful than ASR because the transcription tends to be more accurate.

A rationale of the present study is presented here. The Methods chapter will explain the procedures in greater detail.

Communicative Success between an individual with severe-to-profound hearing loss and an individual with normal hearing was evaluated under the following five conditions:

- Face-to-face-only
- CART-face-to-face
- ASR-face-to-face
- CART-only
- ASR-only

This study was designed to address three basic questions:

1. Text-Transcription Effect

What are the effects of a text-supplement on face-to-face communication?

Hypothesis 1. Communication will be more successful in the ***CART-face-to-face*** *than in the* ***face-to-face-only*** *condition.*

It was hypothesized that the addition of a text supplement (CART) would improve Communicative Success because the deaf participant would be able to *review* the hearing participants utterances.

2. Transcription Method Effect

Does the method of transcription affect Communicative Success?

Hypothesis 2: Communication will be more successful in the ***CART-face-to-face*** *and the* ***CART-only*** *condition than in the* ***ASR-face-to-face*** *and the* ***ASR-only*** *condition.*

Since CART is known to be a highly accurate and well-accepted transcription method, it was predicted to be more successful than ASR.

3. Face-to-Face Effect

How does text-only communication compare to face-to-face-plus-text communication?

Hypothesis 3: Communication will be more successful in the ***CART-face-to-face*** *and* ***ASR-face-to-face*** *conditions than in the* ***CART-only*** *and* ***ASR-only*** *conditions.*

It was predicted that text-plus-face-to-face conditions would be more successful than text-only conditions. Text-only communication is similar to audio-only (telephone) communication for people with normal hearing. Therefore text-only communication lacks certain features such as *visibility* and *simultaneity*.

Communicative Success

Based on previous studies, the following measures of Communicative Success were used to test the hypotheses:

- Task Outcome = performance Map-Task

Previous research on the Map Task (e.g. Doherty-Sneddon et al., 1997; Newlands et al., 1998) did not find that performance on the Map Task differed with communication condition.

- Completion Time

Newlands et al. (1998) showed that although performance on the Map Task was equivalent in face-to-face and CMC (computer-mediated communication), the time to complete the Map Task was significantly longer in CMC. Time to completion was also likely to be a differentiating factor in this study.

- Check Moves

The studies that performed conversational a games analysis found that percentage of Check Moves was the move most likely to differ across communication conditions. Doherty-Sneddon et al. (1997) found that participants used more Check Moves in audio-only and VMC as compared to face-to-face communication. Therefore Check Moves were measured in this study.

- Participant Evaluation

Participant Evaluation was not measured in the earlier Map Task studies. However, it was felt to be a worthwhile measure for this study in order to get a sense, not only of how participants performed, but how they felt about that performance. Participant Evaluation measures could provide insight into communicative effort that may be revealed by the aforementioned measures.

In the next chapter the method for data collection and data analysis are described in detail.

Method

Participants

Data was collected at Gallaudet University in Washington, D.C. from six pairs of conversation partners. One member of each pair had a severe-to-profound hearing loss and the other member had normal hearing. Discourse analysis, by its nature is extremely time consuming, but results in large amounts of data per participants. Therefore a relatively small sample size was used in this study.

Participants with hearing loss[7] were drawn mainly, bit not exclusively from the late-deafened population, i.e. adults with severe-to-profound hearing loss who lost their hearing either progressively or through sudden onset. This subset of the deaf population was chosen because they were deemed the group likely to receive the greatest benefit from ASR-assisted communication. Late-deafened adults tend to prefer a real-time transcription method such as CART to sign-language interpretation because they have attained proficiency in spoken and written English before the onset of deafness. In addition, their speech is intelligible and they are accustomed to producing speech for everyday communication. Participants with hearing loss were recruited from the Washington, D.C. chapter of the Association of Late-Deafened Adults (ALDA) and the Gallaudet University campus. All deaf participants produced speech that was understandable for face-to-face communication. Three men and three women ranging in age from 25 – 64 years old participated in the study. Four were late-deafened with hearing loss onset after age 21. One participant (3D) was prelingually deaf and another

---

[7] A note on terminology: for ease of writing, I will refer to *participants with hearing loss* as **deaf participants** and *participants with normal hearing* as **hearing participants**.

(1D) began to lose her hearing at age 3. However both these participants produced understandable speech and used speech for everyday communication. All had bilateral severe to profound hearing loss, the duration of which ranged from 9 to 40 years. Five of the participants reported wearing a hearing aid in at least one ear. All six participants reported using either hearing aids or some type of assistive listening device to help during face-to-face communication. One participant reported relying completely on speech-reading and not at all on listening, with half of the remaining participants relying mostly on speech-reading, and the other half relying equally on speech-reading and listening. Of the five who responded, two participants reported that they were generally satisfied with face-to-face communication, but three were not. One of the five reported satisfaction in business communications, but not in other aspects of daily life. Five of the six participants were at least somewhat familiar with ASR, and two reported prior experience using it. Formal tests of reading comprehension or other linguistic abilities were not conducted. Table 4 summarizes the demographics for deaf participants.

ID	Gender	Education	current age	age HL onset	years HL	HL Severity Left ear	Right ear	HA's	# years HA use
1D	F	PhD	43	3	40	Profound	Profound	Both	40
2D	F	PhD	61	52	9	Severe	Severe	RE	6
3D	M	BA	25	0	25	Profound	Profound	No	0
4D	M	HS	44	33	11	Severe	Profound	LE	11
5D	M	HS	56	21	35	Profound	Profound	RE	35
6D	F	HS	64	45	19	Profound	Profound	No	0
		mean	*48.8*	*25.7*	*23.2*				
		s.d.	*14.5*	*21.5*	*12.6*				

HL = hearing loss; HA = hearing aid

Table 4: Deaf Participant Demographics

Participants with normal hearing were recruited from graduate programs at Gallaudet University. They had experience communicating with deaf and hard-of-hearing people and working in a PC/Windows computer environment, but no prior experience with ASR. Two men and four women ranging in age from 22 – 60 years old participated in the study. All reported having normal hearing. Five of the six participants were graduate students at Gallaudet University, and all six were experienced in communication with deaf adults using a variety of communication methods.

Participants were informed that this study was not evaluating participants' ability to communicate, but rather ASR's usefulness as an aid to communication (Appendix 1: Informed Consent Form). Participants were matched based on their availability for scheduling experimental sessions.

Participants communicated under five conditions: face-to-face-plus-optional-typing, face-to-face-plus-ASR, face-to-face-plus-CART, ASR-only, and CART-only. In all conditions, all participants communicated by speaking. Participants with normal hearing received information through audition and participants with hearing loss received information via speech-reading and residual hearing and/or text output (Figure 1).

Figure 1: Data collection
(from pilot study)

Training and Assessment

Participants were seen separately for training and assessment. **Participants with hearing loss** were asked to complete a *Technology Use Questionnaire* (Appendix 2) and bring a copy of their most recent audiograms to the first session. Receptive communication abilities (audio-only, video-only, and audio-plus-video) were evaluated using the *CUNY Sentence Test* (Boothroyd et al., 1988). The *CUNY Sentence Test* consists of a series of recorded sentences, presented on a video monitor. Each person was given a sheet with 12 every-day topics. A blank line followed each topic heading. Participants were asked to watch the talker say one sentence, and then to write down what he or she thought the talker said. Participants were tested in audio-only, video-only and audio-plus-video conditions. Results of the CUNY Sentence Test are shown in Table 5.

Deaf Participant	current age	age HL onset	HL Severity	audio	video	audio + video
1D	43	3	Profound	24%	81%	**95%**
2D	61	52	Severe	8%	59%	**97%**
3D	25	0	Profound	5%	66%	**75%**
4D	44	33	Severe	29%	25%	**93%**
5D	56	21	Profound	1%	13%	**44%**
6D	64	45	Profound	0%	35%	**45%**

Table 5: CUNY Sentence Test results

The audio signal was presented in sound field instead of via headphones for face validity purposes: i.e. to be similar to the way speech was presented in the experimental conditions. The audio-only and video-only conditions provide information about the deaf participants' *functional* communication. The audio-only results show which participants had no hearing from a functional point of view. The video-only scores show the deaf participants' speech-reading ability when no audio is presented. It is interesting to note that the participants with the earliest onset of hearing loss (1D and 3D) had the highest video-only receptive communication scores (81% and 66%, respectively). In addition, participants with the lowest audio and video scores (5D and 6D) also had the lowest audio-only scores. For the final portion of their training, participants with hearing loss also had a practice ASR-assisted conversation with a normal-hearing member of the research team.

The six **participants with normal hearing** received two hours of training with Dragon Naturally Speaking, Version 4.0 (the most recent version at the time of data collection). This length of training was chosen as representative of what might be reasonable in a workplace situation, where a co-worker would presumably spend a few hours on the training task but not much longer than this. The participants first "enrolled" or trained the system by reading passages according to the software's training procedures. They did additional training by reading a second story into the system. Participants were also instructed in the basics of ASR spoken commands for sentence management (*"period," "question mark," "new line,"* and *"new paragraph"*) and spoken commands to control the microphone (*"wake up"* and *"go to sleep"*). They learned keyboard and voice commands for error correction and a few short cuts. In addition, they trained the ASR on names of map landmarks that were used for the Map Task (described in next section). Following these procedures, they had approximately thirty minutes of practice in dictating and using commands. Finally, they dictated a written document that was used to assess ASR Transcription Accuracy for each participant's speech under optimal, dictation, conditions (Appendix 3). The transcription accuracy scores for ASR-Dictation are reported in Table 6.

Hearing Participant	ASR Dictation Scores
1	96%
2	89%
3	86%
4	89%
5	91%
6	80%

Table 6: ASR Dictation Transcription Accuracy Scores

Data collection

All sessions were videotaped. Two video cameras were used in a split-screen technique to record each participant's face. The normal hearing participant's voice was recorded with the ASR microphone and the deaf participant wore a lapel microphone for voice recording. A CART reporter was present to make a first-pass transcription of all interactions. In addition, graphical movies of the ASR output and text files of the ASR, CART, and keyboard output were saved (Figures 2 and 3).

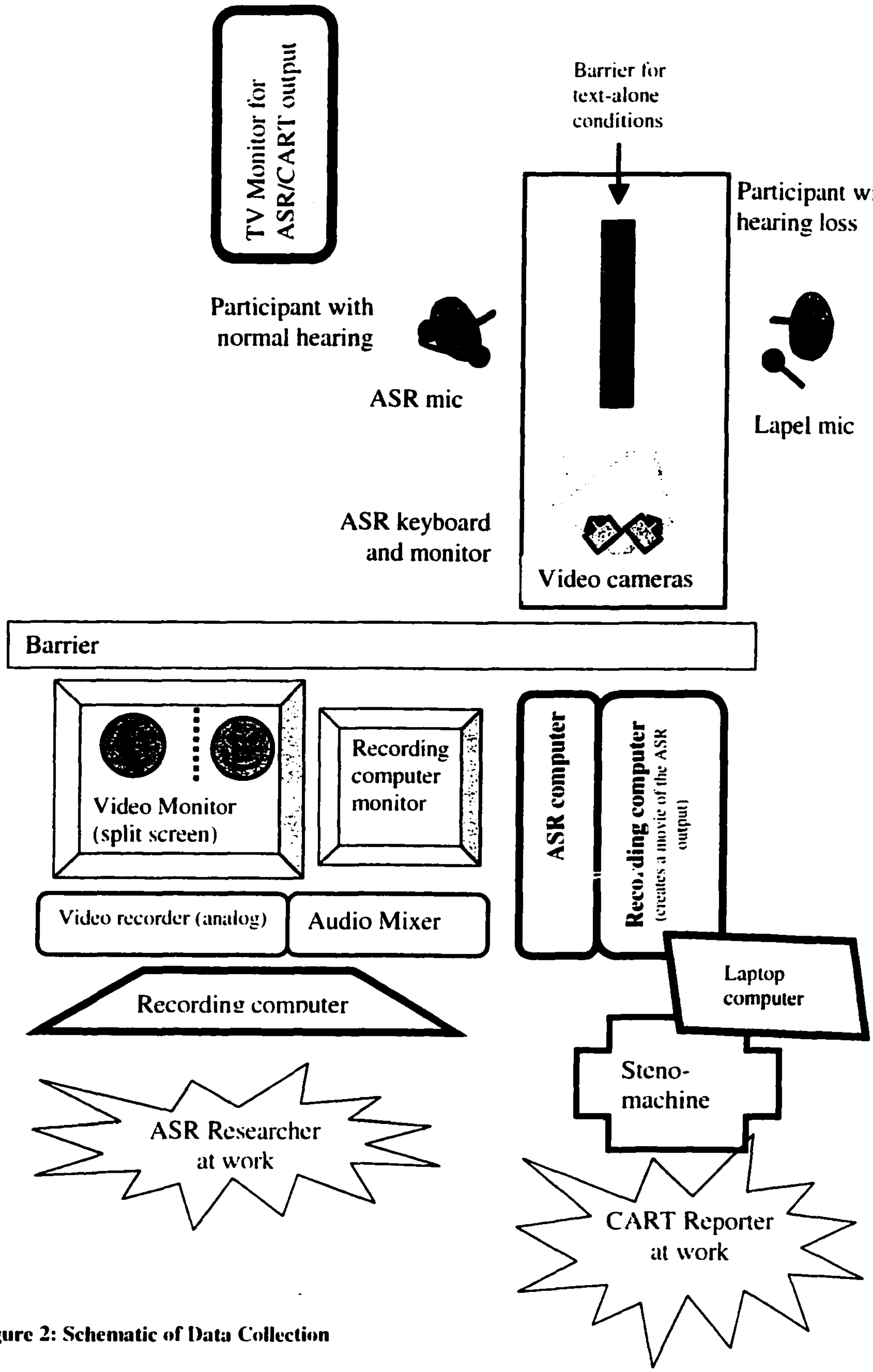


Figure 2: Schematic of Data Collection

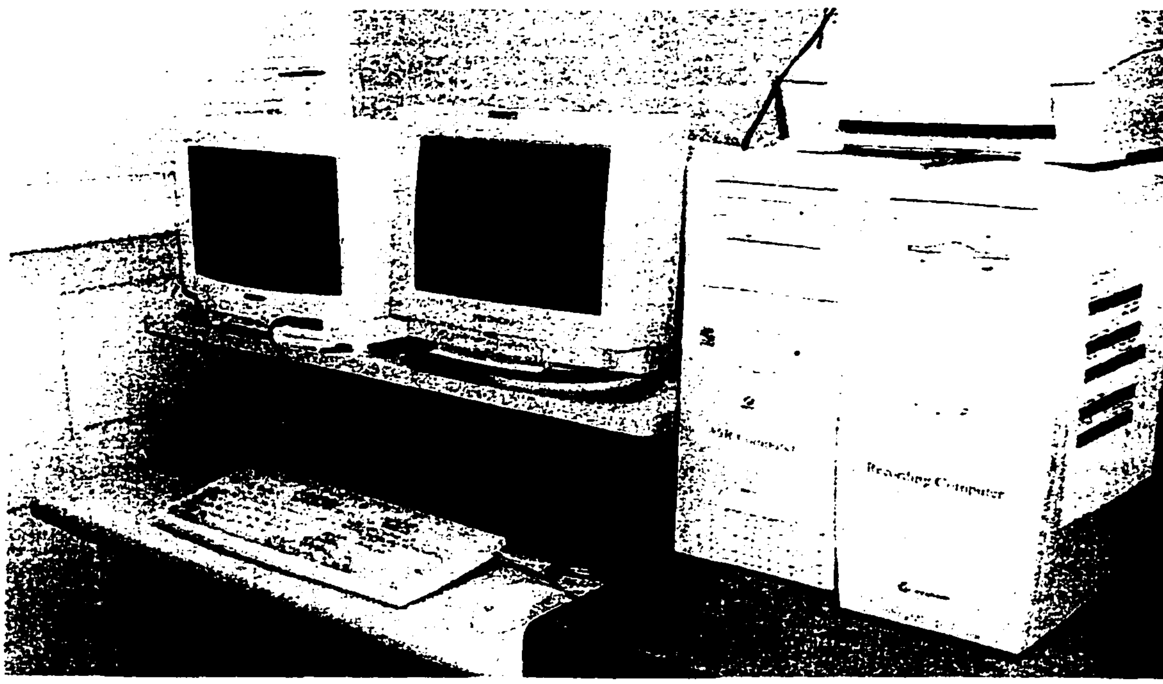


Figure 3: Instrumentation

Experimental Sessions

Each pair of participants met for two experimental sessions. In the first session, they engaged in general conversation for roughly an hour in total. In the second session, the subjects performed a series of collaborative tasks (the Map Task). Thus, two common workplace scenarios were simulated. In both sessions, the order of the conditions was pseudo-randomized. (Please note that the data from the conversation sessions was not analyzed as part of this dissertation, but these sessions served as further practice and familiarization of the participants with each other.)

Conversation Experiment

In the first experimental session, participants were asked to converse in six different communication conditions, each lasting ten minutes. To stimulate discussion, they were provided with a list of topics of general interest and a copy of the *Washington Post*. The communication conditions were:

1. Face-to-face with the ability to use the keyboard if needed to clarify communication
2. Face-to-face using the ASR system
3. Face-to-face using a CART system
4. ASR system only (with a barrier between participants, precluding speech-reading)
5. CART system only (with a barrier between participants, precluding speech-reading)

The latter two conditions were included to simulate a remote/telecommunications situation. Table 7 illustrates the pseudo-randomized testing order. The first condition was a face-to-face-only practice session.

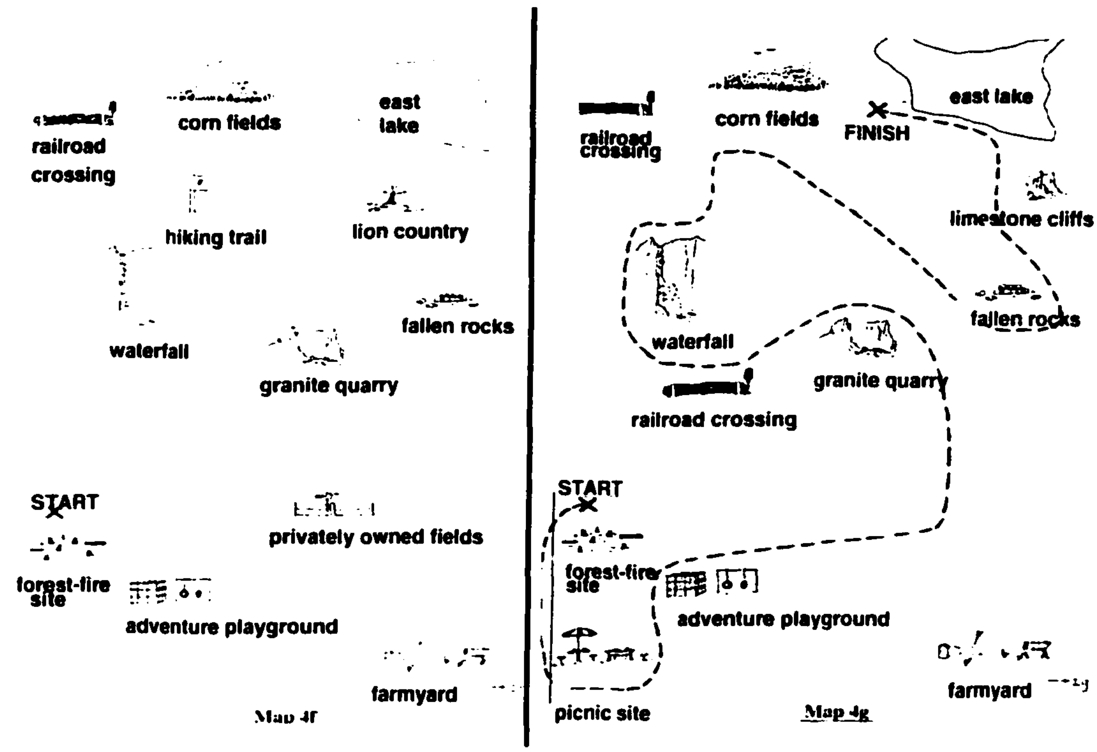
Conversations	**Pair**						
Test Order	**1**	**2**	**3**	**4**	**5**	**6**	
1st	F	F	F	F	F	F	**F = face-to-face**
2nd	CF	A	CF	C	C	C	**FT = face-to-face + typing**
3rd	C	CF	A	CF	FT	FT	**A = ASR-only**
4th	AF	FT	AF	FT	CF	AF	**AF = ASR-face-to-face**
5th	FT	AF	FT	A	AF	A	**C = CART-only**
6th	A	C	C	AF	A	CF	**CF = CART- face-to-face**

Table 7: Conversation Testing Order

After each segment, participants completed feedback questionnaires about their opinions of the experience (Appendices 8-9). Total session time was about two hours.

Map Task Experiment

In the final session, participants completed the Map Task (Brown et al., 1984), a collaborative problem-solving task with a clear goal and a well-defined measure of Communicative Success (Figure 4). Each participant in a pair was given a simple schematic map of the same location, including landmarks. The hearing participant's map had a route drawn on it, but the deaf participant's map had no route drawn. Participants were informed that several landmarks differed slightly in location between the two maps. The participants collaborated to reproduce the route on the deaf person's map.



Map for the participant with hearing loss Map for the participant normal hearing

Figure 4: The Map Task

This goal could only be achieved through the successful exchange of information between participants. The measure of Communicative Success was the extent to which the deaf participant's route matched the original route. One map per condition was solved, with a fifteen-minute time limit given for each condition.

The materials consisted of six maps designed to be of equal communicative difficulty (Anderson et al., 1991). See Appendix 22 for a reproduction of map stimuli. All landmarks were portrayed as line drawings and labeled (e.g. 'old temple', 'banana tree'). Each pair of map routes had the same starting point marked on both maps. The ending point was marked only on the hearing participant's map. Landmarks along the route appeared in the same position on both maps, appeared only on one map, or appeared on both maps, but in different locations. As mentioned in the section above, map landmark names were added to the ASR software's vocabulary during the hearing participants' *training* session.

Since the original maps were developed in Scotland, some landmark names were deemed unacceptable for Standard American English usage. Three individuals with substantial training in linguistics (one professor and two doctoral students) jointly rated the landmark names and agreed on name changes for thirty-eight percent (34 out of 90) of the landmark names. For example, '*telephone box*' was changed to '*telephone booth*' and '*fallen cairn*' was changed to '*fallen rocks*'. See Appendix 4 for a complete list of landmark names and changes.

Participants were given instructions on the Map Task in both oral and written formats (Appendices 5 and 6). In addition, participants with normal hearing were given a list of reminder tips for ASR use (Appendix 7.). As illustrated below (Table 8), communication conditions and maps were pseudo-randomized, except for the initial practice session.

	Pair 1		Pair 2		Pair 3		Pair 4		Pair 5		Pair 6	
Test Order	Condition	Map	Condition	Map	Condition	Map	Condition	Map	Condition	Map	Condition	Map
(1st) Practice	CF	0	CF	0	CF	0	CF	0	CF	0	CF	0
2nd	AF	2	FT	2	A	1	CF	3	C	3	FT	4
3rd	FT	1	A	3	CF	3	AF	2	A	2	AF	5
4th	A	3	AF	5	FT	4	A	5	CF	1	C	2
5th	CF	5	C	4	C	5	FT	4	AF	4	CF	1
6th	C	4	CF	1	AF	2	C	1	FT	5	A	3

FT = face-to-face + optional typing; CF = CART-face-to-face; AF = ASR-face-to-face; C = CART-only; A = ASR-only

Table 8: Map Task Testing Order

Practice session and CART-Face-to-Face Condition

A practice session was conducted to allow participant pairs to familiarize themselves with the task. For the practice session, all participants completed Map 0 in the CART--face-to-face condition. The communication mode for the practice session was *CART-face-to-face* because it gave participants access to the most rich and accurate information. This allowed participants to best familiarize themselves with the task. In this session, as in all face-to-face sessions, conversation partners sat facing each other. The maps were displayed on a double-sided easel placed to one side of the participants. Participants were not able to see each other's maps, but participants' faces and shoulders were visible to each other (Figure 5). The CART output was displayed on a 27" monitor. The monitor was placed to the left of the hearing participant, 58" from the deaf participant (Figure 6).

Figure 5: Data Collection Setup 1

Figure 6: Data Collection Setup 2

Face-to-Face-only[8] Condition

The set-up was the same as the practice condition, except that the CART output was not visible. Participants with normal hearing were permitted to use the computer keyboard to type words at their own discretion. Optional typing is quantitatively different from the real-time transcription techniques being evaluated in his study. The decision to allow hearing participants to type was made in order to assure that participant pairs would be able to complete the Map Task. The pros and cons of this decision are considered in the discussion chapter. The computer monitor and keyboard were placed slightly to the hearing participant's right. If the normal hearing participant chose to type, the output (Georgia font, 20 point) was visible to the deaf participant on a larger monitor above and to the left of the hearing participant.

ASR-Face-to-Face Condition

The set-up was the same as the face-to-face condition, except that the ASR technology (Dragon Naturally Speaking) transcribed the hearing person's speech. Dragon Naturally Speaking ran on a Gateway Pentium III 550 MHz computer with 256 MB of RAM, using a Sound Blaster Audio PCI 64D sound card. No other application software was installed. Two monitors were used. A 27" monitor was situated behind and to the left of the hearing person, and a 17" monitor was at the hearing person's right, out of sight of the deaf participant. Both participants were able to comfortably view a monitor, but only the normal hearing participant had control of the computer, via

[8] Although it would be more accurate to refer to this condition as "face-to-face-plus optional typing," for brevity sake the condition will be referred to as "face-to-face-only throughout this document.

keyboard and headset. The ASR output (Georgia font, 20 point) was visible to both participants on the two monitors. Participants with normal hearing were encouraged, but not required, to use the ASR commands. As an alternative, they were permitted to simply press the "*enter key*" to introduce blank lines between one ASR-transcribed utterance and the next.

ASR-only or CART-only Conditions

Set-up was the same as for the face-to-face-plus-ASR condition or plus-CART, respectively, except that a barrier (12" by 24") was placed on a table. The barrier completely blocked the face and torso of the partner. The barrier also functioned as an easel for the Map-Task. See Figure 1 (page 39) for a picture of the barrier (easel) with a map on it. Note that in Figure 1, which shows the ASR-face-to-face condition, the easel was not placed so as to block participants' view of each other. In the ASR-only and CART-only conditions, the easel was placed so as to block participants' view of their partners. Normal hearing participants were still able to hear the deaf participants and deaf participants made use of the text output and their residual hearing, if any. Therefore, the role of residual hearing, especially for participants with severe hearing loss cannot be discounted. The situation was similar to a VCO (voice carry-over) system from the Relay Service.

Participant Evaluation

After each condition, participants answered a *Session Feedback Questionnaire* in which they were asked to give feedback regarding their experience using the system under the various conditions and their willingness to use it on a regular basis (Appendices 8 and 9). Participants also completed a *Final Feedback Questionnaire* at the conclusion of both experiments (Appendices 10-11). They were asked to rank the communication conditions in terms of communicative efficiency and offer suggestions for improvement. Results of the *Final Feedback Questionnaire* are reported in the next chapter.

Scoring and Coding

This section reports how data was scored and coded in preparation for formal statistical analyses and hypothesis testing. Coding was required for the Map TaskScore, Transcription Editing, Transcription Accuracy and Conversational Games Analysis. Inter-rater reliability[9] (IRR) was obtained for the above measures. Eighty percent or higher inter-rater agreement was considered acceptable (Cozby, 1984, p.54). Raters discussed and resolved any coding disagreements.

Map Task Scores

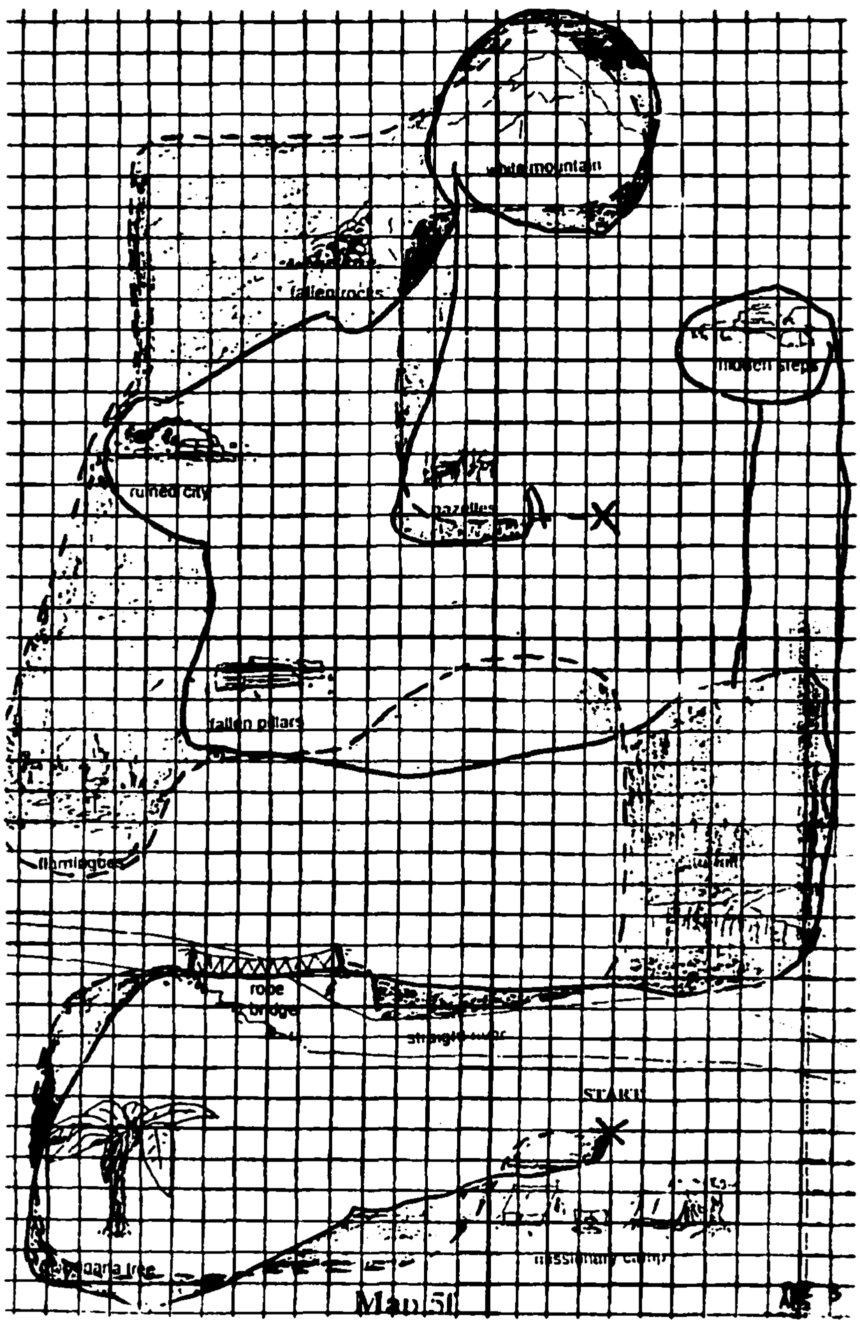
The task outcome measure is based on participant performance on the Map Task.

Deviation Score

Earlier research using the Map-Task scored the maps used a *deviation score,* a method developed by Anderson et al. (1991). With this method, the difference between the collaboratively-drawn route and the original route was calculated. The original route was traced on a one-centimeter square grid of clear transparency paper. The collaboratively-drawn route was then superimposed on the original route and accuracy was assessed by measuring the area of deviation (counting the number of boxes) between the two routes (Figure 7). Partially filled-boxes were included in the tally only if they were more than half full. A high deviation score indicated a less successful performance. These studies found either no significant differences on Map-Task scores between communication conditions (e.g. face-to-face vs. audio-only in Boyle et al., 1994) or a

[9] Note that same researchers distinguish between inter-rater reliability and inter-rater agreement (e.g. Yarrow & Waxler, 1979)

significant learning effect (e.g. computer-mediated communication in Newlands et al. 1998).



The deviation is the area between the original route (solid line) and the collaborative route (dashed line)

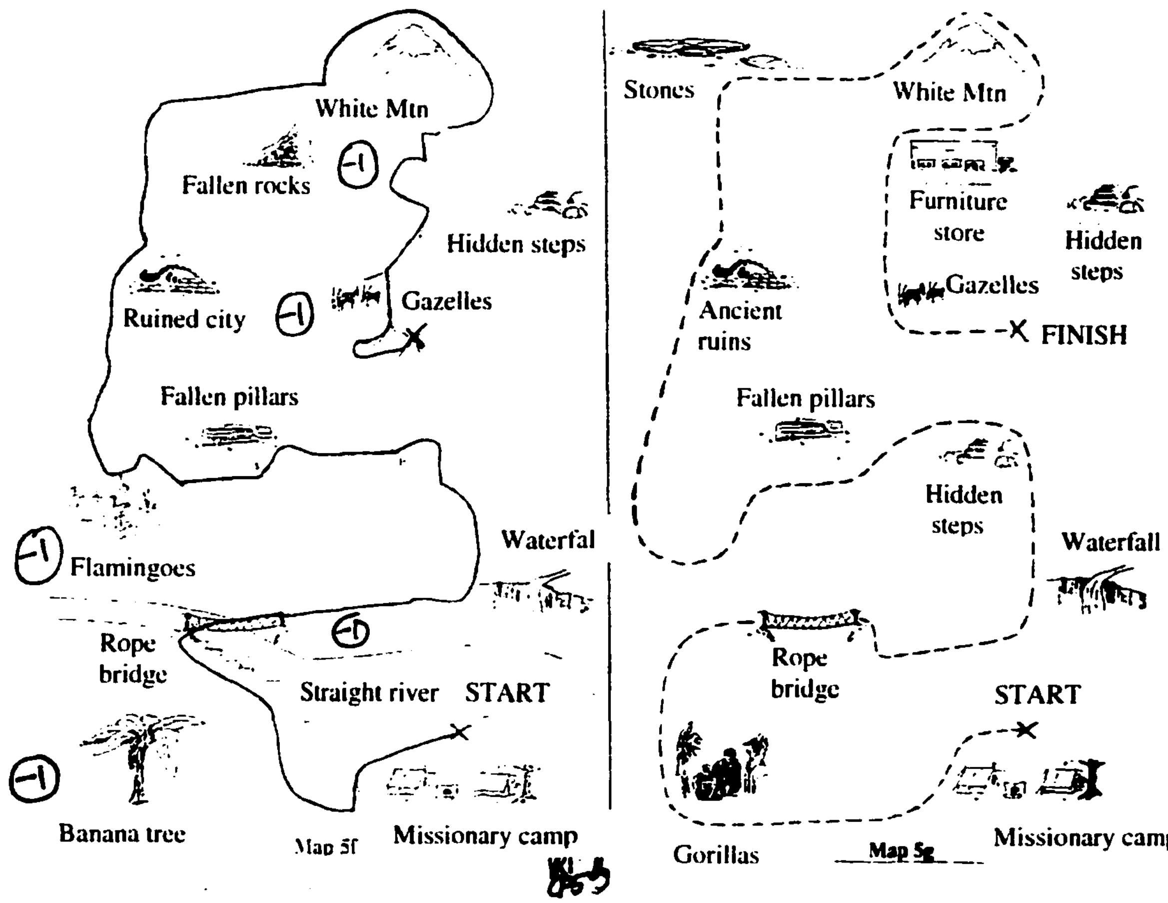
Figure 7: Map deviation score

For this project, pilot analysis using the deviation score revealed no reportable patterns, except a tendency for certain participant pairs to be more successful overall than others.

Landmark Error Score

An alternate method, the *landmark error score,* was developed for this project. This method assessed route reproduction accuracy by counting the number of landmarks on the deaf participant's map that were reached correctly.[10] First, the total number of relevant landmarks on each of the five stimulus maps was tallied. This included the endpoint landmarks and landmarks that appear on only the deaf or hearing participant's map. Then, tallies were made of the number of landmarks not reached successfully on all completed maps. Not reaching a landmark successfully was defined as missing a landmark altogether or drawing the route on the wrong side of a landmark (Figure 8). For details of the scoring procedure, see Appendix 12.

[10] A similar method was developed by Lamb (1996). I discovered this reference after I had already developed the method myself.



Total errors: five out of fourteen landmarks

Figure 8: Landmark error score

Proportion landmark error was calculated by dividing the number of errors by the total number of landmarks on a given map. Proportion errors were then converted to percent correct. On three occasions (Pair 3 Map 1 ASR-only, Pair 4 Map 2 ASR-face-to-face, Pair 4 Map 5 ASR-only), participants did not complete the Map Task in the allotted fifteen minutes. The landmarks not reached were scored as errors (Table 9).

	PAIR				
CONDITION	**1**	**2**	**3**	**4**	**5**
Face-to-Face	64%	64%	62%	85%	29%
CART-FTF	64%	93%	69%	85%	36%
CART	92%	77%	50%	93%	31%
ASR-FTF	64%	64%	50%	**43%***	62%
ASR	69%	77%	**43%***	**64%***	21%

* Pairs that did not finish Map Task in 15 minutes:

Pair 3 ASR-only: 8 landmarks out of total 14 reached. 2 errors. 6 landmarks were not reached at all. Total errors = 8.

Pair 4 ASR-FTF: 9 landmarks out of 14 reached. 0 errors. 5 landmarks were not reached at all. Total errors = 5.

Pair 4 ASR-only: 11 landmarks reached out of 14 reached. 5 landmarks were not reached at all. Total errors = 8

Table 9: Map Scores by condition

A research assistant scored all thirty maps after receiving training from the author. To assure inter-rater reliability (IRR), the author scored five of the thirty maps. The five maps were chosen such that one of each of the five maps was rated. The number of discrepancies between the two raters was tallied. The equation for inter-rater reliability was:

$$IRR = 1 - \frac{\text{total discrepancies}}{\text{total landmarks}}$$

Inter-rater reliability across the five maps was 92.6% (Table 10).

	Landmark Errors			Landmarks	
Dialogue	**Rater A**	**Rater B**	**Discrepancy**	**per Map**	**IRR***
PR5 CF1	11	9	2	14	**85.7%**
PR4 AF2	6	5	1	14	**92.9%**
PR5 C3	8	9	1	13	**92.3%**
PR5 AF4	5	5	0	13	**100.0%**
PR2 AF5	6	5	1	14	**92.9%**
Sums	**36**	**33**	**5**	**68**	**92.6%**

***IRR** = 1-total disagreements/total landmarks

Table 10: Inter-rater Reliability for Map-Task Scores

Transcript editing

A CART reporter transcribed all experimental sessions in real-time, meaning while participants were completing the Map Task. During the CART-only and CART-face-to-face conditions, CART transcription was visible to the deaf participant. During the remaining experimental conditions, the CART transcript was not visible to participants, but was saved to provide a base transcript for later analysis. After data collection, the CART transcripts of each dialogue were edited and compared to the recordings. Hesitations, dysfluencies, and fillers were added. The transcription was also marked to identify areas of overlapping speech and interruptions. A ***Transcription-Editing Manual*** (Appendix 13) was created to describe the rules for identification and coding of discourse units. (See Appendix 14 for a summary of the number of words per dialogue.)

Inter-rater reliability for transcription editing was obtained by comparing two of the thirty dialogues, which were scored by the author and a research assistant. The number of units per dialogue was tallied. Units included, not only words, but also other utterances such as dysfluencies, hesitations and fillers. Then the two transcripts for a particular dialogue were compared and the number of discrepancies (units that the two raters did not agree upon) was tallied. The equation for inter-rater reliability was:

$$IRR = 1 - \frac{\text{total discrepancies}}{\text{total units}}$$

For total units, the smaller total (from the two raters) was used so that the IRR percent score would be lower (smaller denominator yields larger quotient). Inter-rater reliability across the two dialogues was 98.3% (Table 11).

	Units per dialogue			
Dialogue	AH	JE	discrepancies	IRR*
Pr4 AF2	1304	1295	24	98.1%
Pr6 A3	874	870	13	98.5%

***IRR** = 1 - discrepancies/total units per dialogue

Table 11: Inter-rater Reliability for Transcription editing

Optional Typing in the Face-to-Face-only condition

As mentioned in the Methods chapter (p. 50), the Face-to-Face-only condition is more accurately referred to as face-to-face-plus-optional-typing. A brief summary of the frequency of typing incidences is presented here. Although all hearing participants were given the opportunity to type, only two of the six chose to do so. These were participants 5H and 6H whose deaf partners had the lowest Audio-Visual Speech-Perception scores of all the deaf participants. (44 and 45%, respectively). Participant 5H (the hearing participant in pair 5) typed 18.8 % of the words that he produced (76 out of 404). The 404 includes words both spoken and typed. Participant 6H typed 6.8% of words that she produced (24/354). In terms of performance on the Map Task, Pair 5's score of 29% was worse than the average of 62%. Pair 6 performed slightly above average (69%). It could be argued that both pair's performance might have been even worse had they not been permitted to type.

Transcription Accuracy (ASR/CART)

Transcription Accuracy was coded to determine how accurately the ASR and CART systems transcribed the hearing participants speech. The edited transcripts were compared to the real-time transcription produced by the ASR software and the CART reporter. The ASR and CART-generated transcriptions were scored for accuracy according to the protocols from the *NTID Test of ASR Readability* (Stuckless, 1998). ASR and CART transcription errors were coded as substitutions, omissions, and additions. Transcription Accuracy for the Map-Task dialogues in this study had different coding requirements than those for which the *NTID Test* was designed. The *NTID Test* was created to assess ASR Transcription Accuracy for a classroom lecture which is more similar to dictation than conversation. Any deviation from the *NTID* coding method has been reported in an ***Transcription accuracy coding manual*** (Appendix 15). The Transcription Accuracy score was arrived at by calculating the total number of units that the hearing participant uttered:

$$\text{Transcription Accuracy} = 1 - \frac{\text{transcription errors}}{\text{total units}}$$

Some examples of ASR transcription errors follow:

Example 1: ASR transcription error (Pair 5 A2

Hearing Participant: Yes.

ASR: Gets

Example 2: ASR transcription error (Pair 1 AF4)

Hearing Participant: The line moves into the right.

ASR: bowline moves into the right

Example 3: ASR transcription error (Pair 1 AF4)

Hearing Participant: No, at the waterfall, you went north.

ASR: now at the waterfall U.N.'s north

Inter-rater reliability for Transcription Accuracy was obtained by comparing two of the thirty dialogues. These dialogues were scored by the author and the research assistant. The number of transcription errors per dialogue was tallied for each rater. Then the two transcripts for a particular dialogue were compared and the discrepancies (number of transcription errors that the two raters did not agree upon) were tallied. The equation for inter-rater reliability was:

$$IRR = 1 - \frac{\text{total discrepancies}}{\text{total units uttered by the hearing participant}}$$

For total units, the smaller total (from the two raters) was used so that the IRR percent score would be lower (smaller denominator yields larger quotient). Inter-rater reliability across the two dialogues was 98.3% (Table 12).

Transcription Accuracy errors				**Total units**	
Dialogue	AH	JE	discrepancies	**(Hearing participant)**	IRR*
Pr4 AF2	144	149	15	665	97.7%
Pr6 A3	57	59	6	445	98.7%

***IRR** = 1 - discrepancies/ units uttered by the hearing participant

Table 12: Inter-rater Reliability for Transcription accuracy

ASR-Dictation Accuracy

To assess ASR Transcription Accuracy under optimal conditions, hearing participants dictated a 231-word article into the ASR system (Appendix 3). The Transcription Accuracy was scored in a manner similar to Transcription Accuracy for the dialogues. See Appendix 16 for the ASR-Dictation Scoring Rules. The IRR results are reported below (Table 13).

Hearing Participant	Transcription errors				
	EH	JE	Discrepancies	Total words	IRR*
1	9	11	2	231	99.1%
2	21	23	2	231	99.1%
3	32	33	1	231	99.6%
4	19	20	1	231	99.6%
5	19	19	0	231	100%
6	44	45	1	231	99.6%
Total	144	151	7	1386	99.5%

***IRR** = 1 - discrepancies/total words

Table 13: IRR for ASR-Dictation Accuracy

Check Moves Analysis

A *conversational games analysis* (Kowtko et al., 1992) was conducted. This analysis method was originally designed for the Map-Task dialogues. Each utterance was coded as a particular *move.* A *conversational move* is defined as "the perceived conversational function the speaker intends to accomplish with that utterance" (Newlands, 1998, p.157). Each utterance was coded by taking into account the semantic content of the utterance, the location of the utterance within the dialogue, and utterance prosody. Figure 9 illustrates the steps taken to decide the *move* category of an utterance.

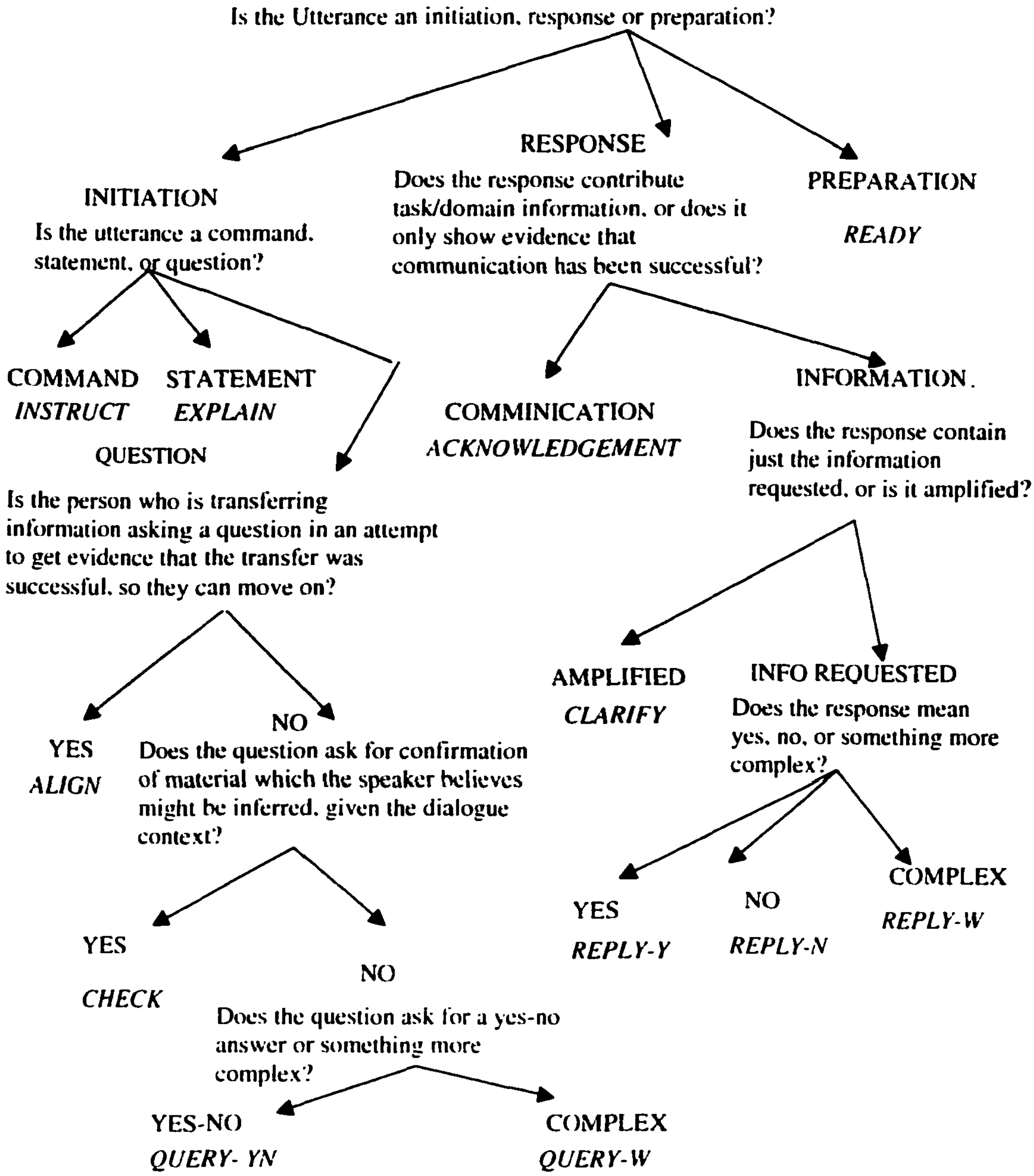


Figure 9: Decision Tree for Conversational Moves.
Reprinted with permission from Carletta et al. 'The Reliability of a Dialogue Structure Coding Scheme', *Computational Linguistics, 23:1* (March, 1997), pp. 13-31, Fig. 1 (p. 15). © 1997 by the Association for Computational Linguistics.

Check Moves were of particular interest to the question of communicative efficiency. These moves are uttered when participants feel the need to confirm mutual understanding. In Table 14, a brief dialogue illustrates a *CHECK* move.

	Utterance	Conversational Move
Hearing Participant:	'Do you have a picnic table?'	QUERY
Deaf Participant:	'A what?'	CHECK
Hearing Participant:	'A picnic table.'	CLARIFY
Deaf Participant:	'Oh. Yes, I do.'	REPLY

Table 14: A dialogue with utterances labeled by conversational move
(adapted from Clark & Brennan, 1991).

The author of this dissertation received training from Gwyneth Doherty-Sneddon of the University of Stirling in Scotland. Dr. Doherty-Sneddon is one of the creators of Conversational Games Analysis and has published extensively the results of experiments using this analysis system. Upon completion of training, Dr. Doherty-Sneddon was satisfied that the author had mastered the theoretical underpinnings and technical aspects of the coding scheme. The author began coding all thirty dialogues. Dr. Doherty-Sneddon was available via e-mail when coding questions arose.

To the extent possible, each turn of Map-Task dialogue was classified as one of the conversational moves in the coding scheme. Less than one percent of turns per dialogue could not be coded as moves within this system. For example, a comment about one of the landmarks (e.g. "That's a funny looking picnic table.") or an utterance directed toward the ASR system (e.g. "new line, scratch that, correct that") were excluded in the tally of total moves. After coding was completed, the moves were tallied. The tallies were converted to a proportion by dividing the total moves for a particular category by the total number of moves, excluding uncoded turns. For illustration purposes, Pair 4 face-to-face-only condition was chosen as a typical example of the tally of conversational moves in a typical Map Task (Table 15) (See Appendix 17 for the **Conversational Games Analysis Manual.**)

Pair 4 Face-to-face	Tally	Percent of total
Acknowledge	29	27%
Reply	24	23%
Instruct	18	17%
Query	14	13%
Check	**11**	**10%**
Explain	7	7%
Align	2	2%
Clarify	1	1%
TOTAL	106	100%

Table 15: Conversational Moves (Pair 4 Face-to-face-only)

Conversational games analysis was conducted on all thirty dialogues (Appendix 18). Inter-rater reliability was calculated for two of the dialogues. These dialogues were scored by the author and the research assistant. The total number of Moves per dialogue was tallied for each rater. Then the two transcripts for a particular dialogue were compared and discrepancies (number of Moves that the two raters did not agree upon) were tallied (Table 16). The equation for inter-rater reliability was:

$$IRR = 1 - \frac{\text{total discrepancies}}{\text{total Moves}}$$

Inter-rater reliability averaged across the two dialogues was 87%.

Dialogue	Total Moves per dialogue AH	JE	Discrepancies	IRR
Pr1 CF1	185	193	19	89.7%
Pr1 A3	137	134	21	84.3%

*IRR = 1 - discrepancies/total Moves

Table 16: Inter-rater Reliability for Total Moves

As will be discussed in the next chapter, Check Moves were the only moves that revealed significant differences, IRR for Check Moves are reported (Table 17). The equation for inter-rater reliability was:

$$IRR = 1 - \frac{\text{total discrepancies}}{\text{total Check Moves}}$$

Inter-rater reliability across the two dialogues was 84.5%.

Dialogue	Check Moves per dialogue AH	JE	discrepancies	IRR
Pr1 CF1	14	14	2	85.7%
Pr1 A3	12	12	2	83.3%

*IRR = 1 - discrepancies/total Check Moves

Table 17: Inter-rater Reliability for Check Moves

The IRR for conversational games analysis are comparable to the IRR levels reported by Doherty-Sneddon et al. (1997) and Newlands (1998).

The results of hypothesis testing are presented in the next chapter. Note that a summary of all the raw data can be found in Appendix 19.

Results

As mentioned above, the following research hypotheses were addressed:

Hypothesis 1. Communication will be more successful in the ***CART-face-to-face*** *than in the* ***face-to-face-only*** *condition (Text-Transcription Effect).*

Hypothesis 2: Communication will be more successful in the ***CART-face-to-face*** *and the* ***CART-only*** *condition than in the* ***ASR-face-to-face*** *and the* ***ASR-only*** *condition (Transcription Method Effect).*

Hypothesis 3: Communication will be more successful in the ***CART-face-to-face*** *and* ***ASR-face-to-face*** *conditions than in the* ***CART-only*** *and* ***ASR-only*** *conditions (Face-to-Face Effect).*

Rationale for Statistical Design

The hypotheses were addressed in terms of four measures of Communicative Success: Task Outcome, Time to Completion, Check Moves and Participant Evaluation. The decision was made to look at specific contrasts of the three hypotheses by means of two-way ANOVA's. This approach was adopted rather than a global three-way ANOVA because there was some evidence that the error variance varied for the different ANOVA's.

In addition to addressing the hypotheses in terms of the above-mentioned measures, post-hoc analyses were performed on contrasts subsequently deemed of interest. Analyses of Transcription Accuracy and Audio-Visual Speech Perception were also conducted.

Task Outcome Results

The Task Outcome measure is based on participant performance on the Map Task. The Map-Task score is arrived at by dividing the number of landmarks reached correctly by the total landmarks on a given map. Each of the hypotheses with respect to this measure was tested using a repeated-measures ANOVA. Performing further statistics on proportions and percentages is problematic because, like regression, analysis of variance assumes that error variance is the same for all conditions (homogeneity of variance assumption). This is not true for proportion data. The error variance of a measured proportion is at maximum when $\underline{p} = 0.5$ and at its minimum (0.0) when $\underline{p} = 0.0$ or 1.0. The arcsine transform makes the error variance approximately equal for all values of $\underline{p}$ (except for $\underline{p} = 0.0$ or 1.0). All proportion or percentage data were subjected to the arsine transform for purposes of statistical analysis[11] and then transformed back to proportions (percentages) for reporting mean values. In addition, the standard error (s.e.) of each mean was estimated assuming the binomial probability distribution for each observed percentage score.

Text-Transcription Effect

Hypothesis 1: Map-task scores will be higher in the CART-face-to-face condition than in the face-to-face-only condition.

A two-way repeated-measures ANOVA was run to compare completion times in the CART-face-to-face and face-to-face-only conditions. Table 1 summarizes the results

[11] (ASIN((SQRT(X))/ASIN(1)) for proportions and (100*ASIN((SQRT(X/100)))/ASIN(1)) for percentages.

for the repeated measures ANOVA for the Map Task. Bear in mind that the sums of squares shown in the table are in arcsine units. The main effect for condition (C) was not significant (Table 18, Figure 10). Therefore Hypothesis 1 was not supported for the Task Outcome scores.

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	SIGNIF.
Condition (C)	0.02	1	0.02	0.26	0.63
Participant Pairs (P)	1.64	5	0.32	3.88	0.08
CP	0.42	5	0.08		
TOTAL	2.08	11			

Table 18: Task Outcome ANOVA table for Text-Transcription Effect

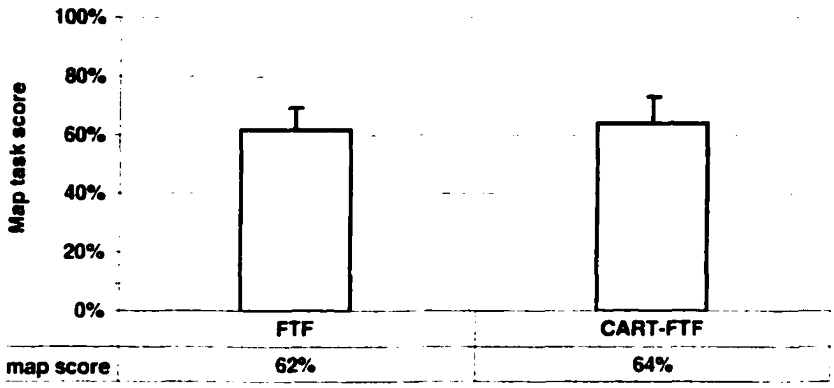


Figure 10: Task Outcome for Text-Transcription Effect
(means and s.e.)

Transcription Method and Face-to-Face Effects

Hypothesis 2: Map-task scores will be higher in CART-face-to-face and CART-only than ASR-face-to-face and ASR-only.

Hypothesis 3: Map-task scores will be higher in CART-face-to-face and ASR-face-to-face than CART-only and ASR-only.

Hypotheses 2 and 3 were tested using a three-way repeated-measures ANOVA which compared all four text-based communication conditions. The main effects for both Transcription Method (Hypothesis 2) and addition of Face-to-Face Effect (Hypothesis 3) were not significant (Table 19, Figures 11 and 12). Thus Hypotheses 2 and 3 were not supported for the Task Outcome scores.

SOURCE OF VARIATION	SUMS OF SQUARES	dF	MEAN SQUARES	F	SIGNIF.
Transcription Method (M)	**0.47**	1	**0.47**	**4.29**	**0.09**[12]
FTF cues (F)	**0.01**	1	**0.01**	**0.11**	**0.75**
MF	0.014	1	0.01	0.17	0.69
Participant Pairs (P)	2.82	5	0.56	6.70	0.03
FP	0.62	5	0.12	1.47	0.34
MP	0.55	5	0.11	1.31	0.39
MFP	0.42	5	0.08		
TOTAL	4.92	23			

Table 19: Task Outcome ANOVA table for hypotheses 2 and 3

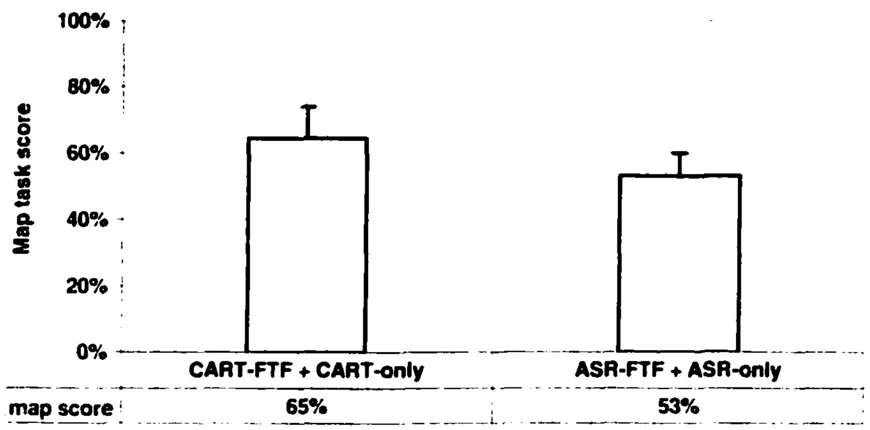


Figure 11: Task Outcome for Transcription Method effect
(means and s.e.)

[12] For Transcription Method, a fixed effects ANOVA brought the significance level to p = 0.06 and F = 5.6.

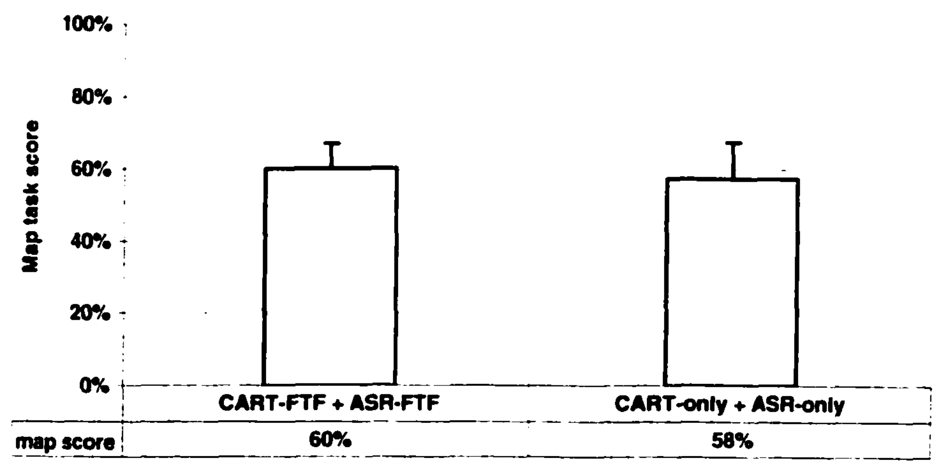


Figure 12: Task Outcome for Face-to-Face Effect
(means and s.e.)

In summary, there were no significant differences among the five communication conditions in terms of Task Outcome. On the next page, Figures 13 summarizes the mean Task Outcome scores for the five communication conditions. Figure 14 shows map scores for individual participants.

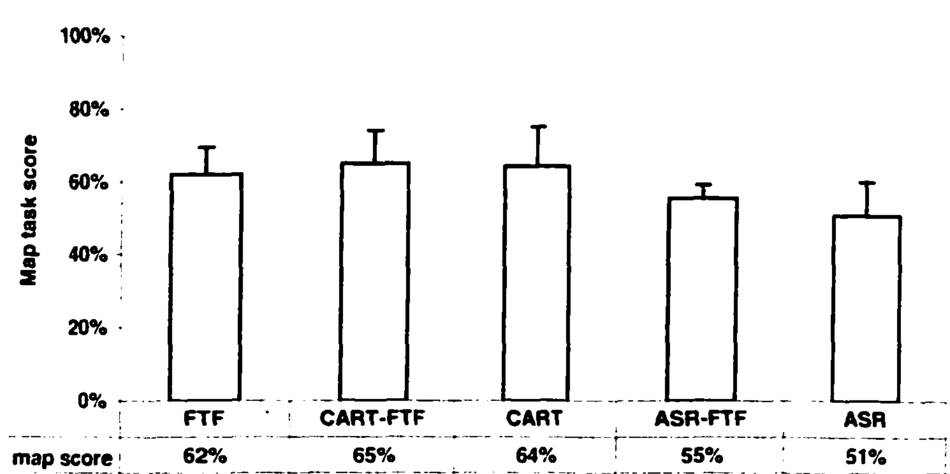


Figure 13: Summary of Task Outcome for all communication conditions
(means and s.e.)

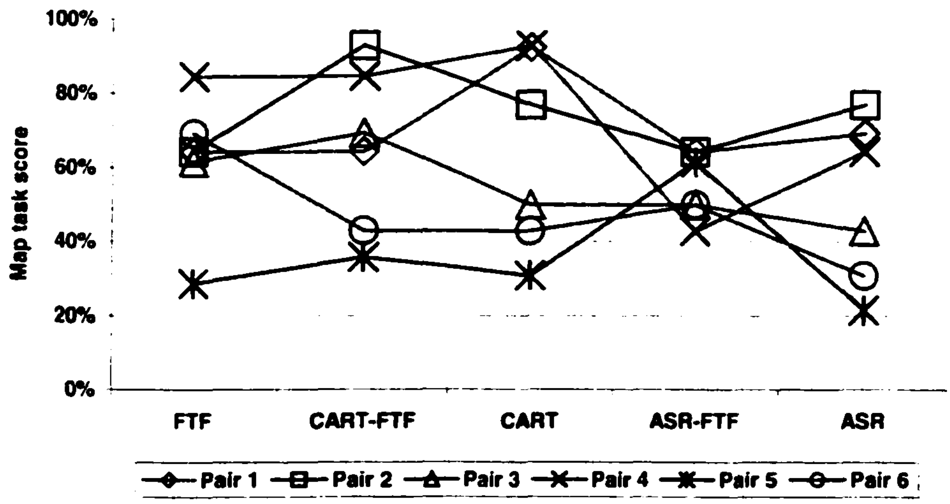


Figure 14: Map-task scores for individual participant pairs

Although this measure of Communicative Success (the Map-Task score) did not differentiate among the communication conditions, some of the other measures did show differences, as we will see in the following sections.

Completion Time results

Time to complete the Map Task was measured from the video recordings. Maximum time permitted to complete the Map Task was fifteen minutes. Shorter Completion Time was considered an indication of greater communicative efficiency. On four occasions, participants had not completed the maps in the time allotted. Pair 5, ASR-face-to-face, were permitted to go over by 30 sec. In the other situations, pair 3 (ASR-alone) and pair 4 (ASR-face-to-face and ASR-alone), participants were asked to stop and Completion Time was recorded as fifteen minutes.

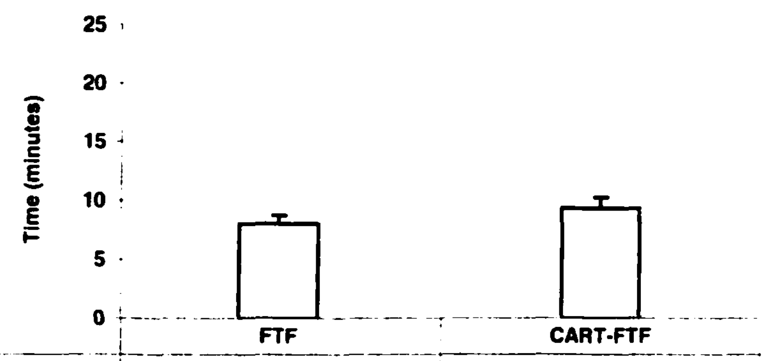
Text-Transcription Effect

Hypothesis 1: CART-face-to-face will require less time than face-to-face-only.

A two-way repeated-measures ANOVA was run to compare completion times in the CART-face-to-face and face-to-face-only conditions. Both of the main effects, condition and participant pair, were not significant (Table 20, Figure 15). Therefore Hypothesis 1 was not supported for completion times.

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	SIGNIF.
Condition (C)	4.94	1	4.94	1.11	0.34
Participant Pairs (P)	16.57	5	3.31	0.74	0.62
CP	22.29	5	4.46		
TOTAL	43.81	11			

Table 20: Completion Time ANOVA table for Text-Transcription Effect



	FTF	CART-FTF
mean time	8.05	9.34

Figure 15: Completion Time for Text-Transcription Effect
(means and s.e.)

Transcription Method and Face-to-Face Effects

Hypothesis 2: CART-face-to-face and CART-only will require less time than ASR-face-to-face and ASR-only.

Hypothesis 3: CART-face-to-face and ASR-face-to-face will require less time than CART-only and ASR-only.

Hypotheses 2 and 3 were tested using a three-way repeated-measures ANOVA which compared all four text-based communication conditions (Table 21, Figure 16). The main effect for Transcription Method (Hypothesis 2) was significant ($p = 0.045$). Thus Hypothesis 2 was accepted since Time to Completion in the CART conditions was significantly less than that for the ASR conditions. Note that this finding was significant even though participants were required to stop interacting after fifteen minutes.

SOURCE OF VARIATION	SUMS OF SQUARES	dF	MEAN SQUARES	F	SIGNIF.
Transcription Method (M)	59.85	1	59.85	6.99	0.045*
FTF cues (F)	3.15	1	3.15	0.55	0.50
MF	5.70	1	5.70	1.99	0.22
Participant Pairs (P)	79.88	5	15.96	5.57	0.04
FP	28.26	5	5.65	1.97	0.23
MP	42.78	5	8.56	2.99	0.13
MFP	14.33	5	2.87		
TOTAL	233.93	23			

Significant at $p < 0.05$

Table 21: Completion Time ANOVA table for Hypotheses 2 and 3

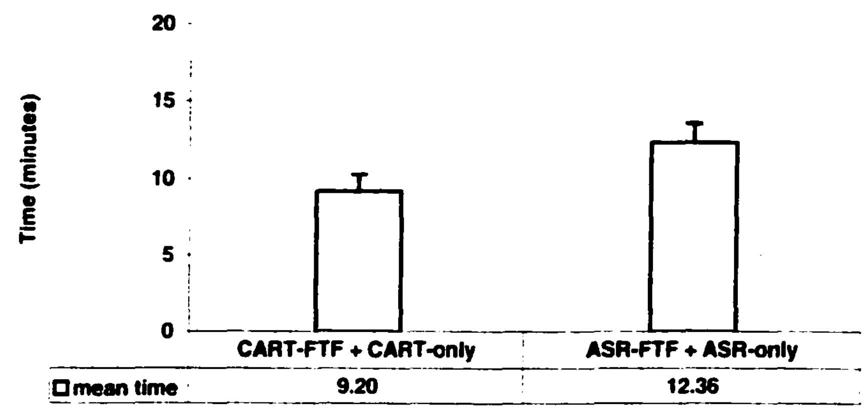


Figure 16: Completion Time for Transcription Method Effect
(means and s.e.)

The main effect for addition of face-to-face cues (Hypothesis 3) was not significant (Table 21, Figure 17). Thus Hypothesis 3 was not supported since Completion Time to in the text-only conditions was not significantly different from the text-plus-face-to-face conditions.

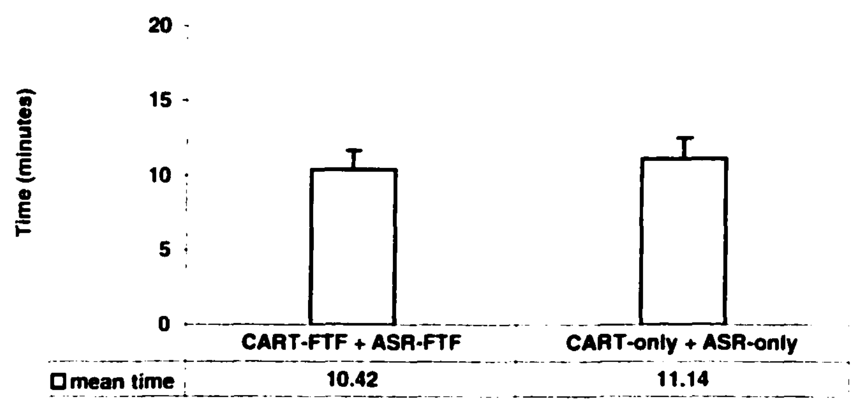


Figure 17: Completion Time to for Face-to-Face Effect
(means and s.e.)

In summary, the results indicate that CART was more efficient than ASR with respect to task completion times, but that other effects, addition of text to face-to-face and addition of face-to-face cues to text, did not have a significant effect on Completion

Times. Figure 18 summarizes the results of hypothesis testing across all five conditions[13]. Figure 19 shows completion times for individual participant pairs.

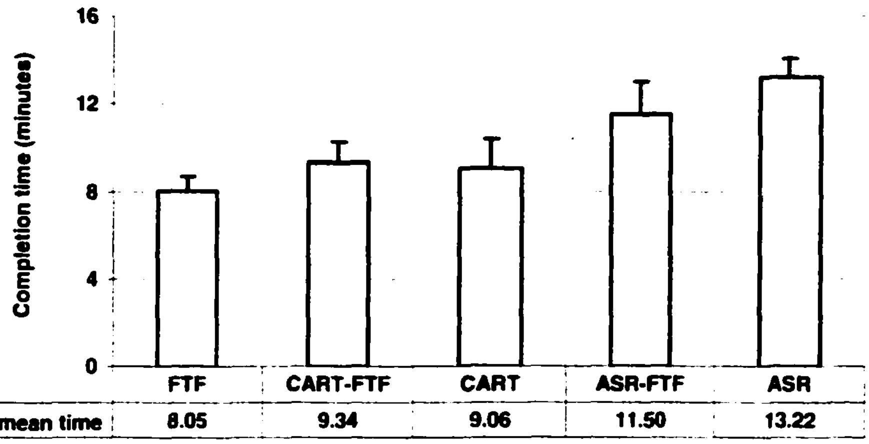


Figure 18: Summary of mean completion times
(means and s.e.)

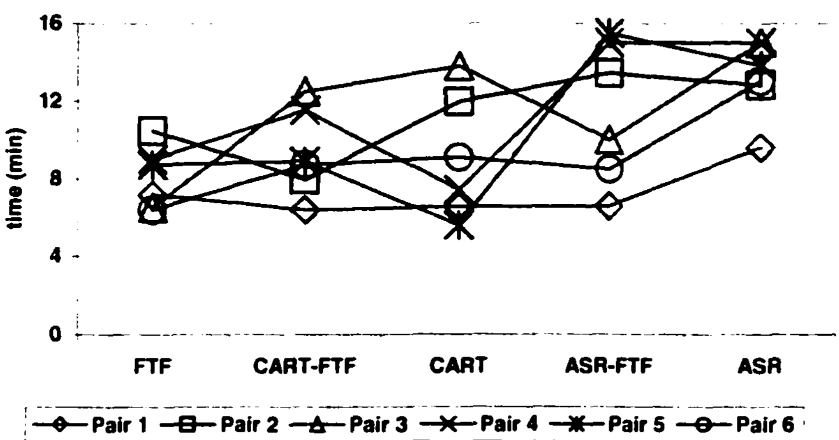


Figure 19: Completion times for individual participant pairs

[13] A post-hoc ANOVA of Completion Time did not find significant differences between ASR-FTF and ASR-only ($F = 3.37$, $p = 0.1$).

Check Move results

Recall from the methods chapter that each of the Map-Task dialogues underwent a conversational games analysis (Kowtko et al., 1992). Check Moves were of particular interest because earlier research (e.g. Doherty-Sneddon et al. 1997) found that increased Check Moves were associated with Map-Task dialogues that were more effortful.[14] The percentage of Check Moves was obtained by dividing the number of Check Moves by the total number of all Moves in a given dialogue. Since the Check Move scores were in the form of percentages, the arcsine transform was used to stabilize the error variance (as was done for the Map-Task scores).

<u>Text-Transcription Effect</u>

Hypothesis 1: Participants will produce fewer Check Moves in the CART-face-to-face condition than in the face-to-face-only condition.

A two-way repeated-measures ANOVA was run to compare Check Moves in the CART-face-to-face and face-to-face-only conditions. The main effect for condition was significant ($p = 0.04$) (Table 22, Figure 20). Therefore hypothesis 1 was accepted for Check Moves.

[14] Although analyses were run to compare all Move-types, Check Moves were the only type to results in significant differences. These results are reported here.

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	SIGNIF.
Condition (C)	0.13	1	0.13	7.06	0.04
Participant Pairs (P)	0.04	5	0.01	0.46	0.79
CP	0.09	5	0.02		
TOTAL	0.25	11			

Table 22: Check move ANOVA table for Text-Transcription Effect

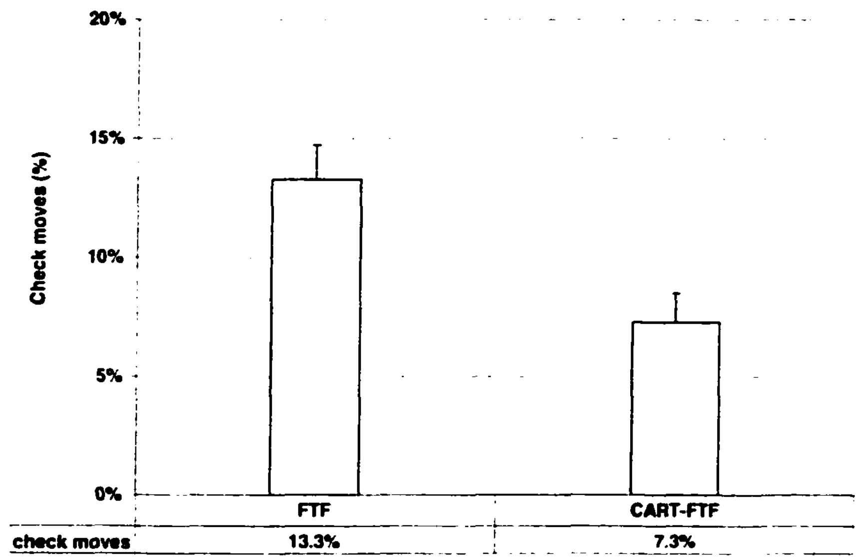


Figure 20: Check Moves for Text-Transcription Effect
(means and s.e.)

<u>Transcription Method and Face-to-Face Effect</u>

Hypothesis 2: Participants will produce fewer Check Moves in CART-face-to-face and CART-only than ASR-face-to-face and ASR-only.

Hypothesis 3: Participants will produce fewer Check Moves in CART-face-to-face and ASR-face-to-face than CART-only and ASR-only.

Hypotheses 2 and 3 were tested using a three-way repeated-measures ANOVA which compared all four text-based communication conditions (Table 23, Figures 21 and

22). The main effects for both Transcription Method (Hypothesis 2) and addition of face-to-face cues (Hypothesis 3) were not significant. Thus Hypotheses 2 and 3 were not supported for Check Moves.

SOURCE OF VARIATION	SUMS OF SQUARES	dF	MEAN SQUARES	F	SIGNIF.
Transcription Method (M)	0.004	1	0.004	0.72	0.56
FTF cues (F)	0.03	1	0.03	0.63	0.53
MF	0.02	1	0.02	1.613	0.26
Participant Pairs (P)	0.05	5	0.01	0.69	0.65
FP	0.03	5	0.006	0.41	0.82
MP	0.21	5	0.04	2.94	0.13
MFP	0.074	5	0.01		
TOTAL	0.43	23			

Table 23: Check move ANOVA table for Hypotheses 2 and 3

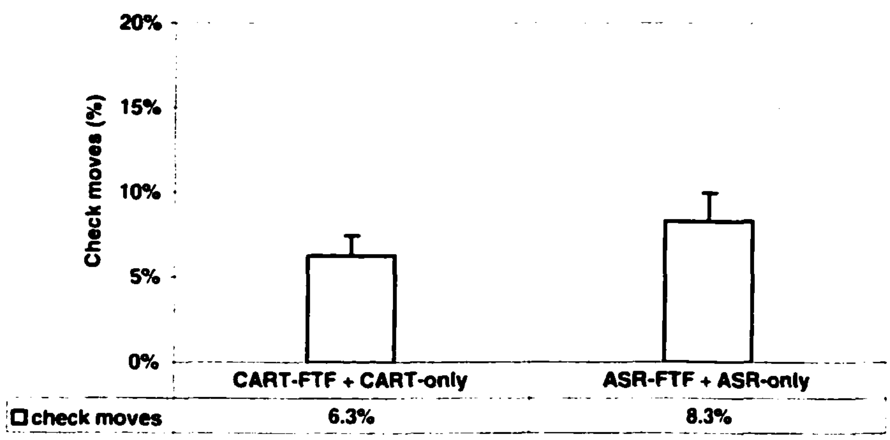


Figure 21: Check Moves for Transcription Method Effect
(mean and s.e.)

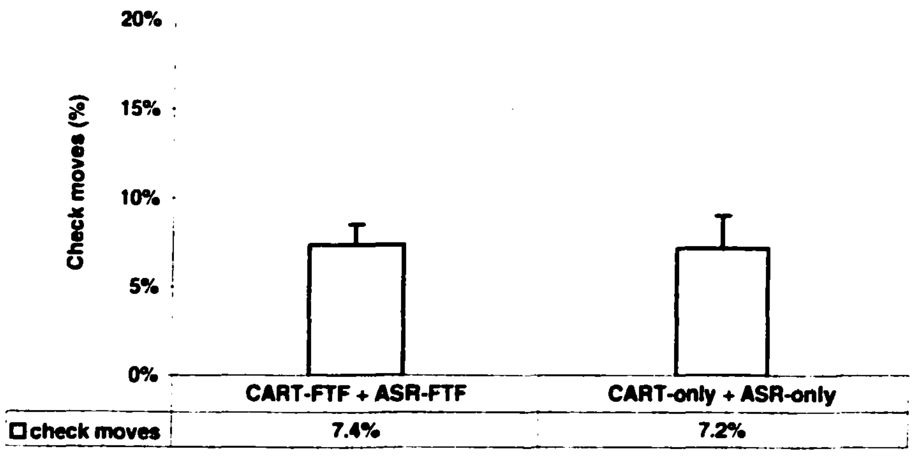


Figure 22: Check Moves for Face-to-Face Effect
(mean and s.e.)

In summary, the results indicate that the CART-Face-to-Face condition was more efficient than the Face-to-Face-only condition with respect to Check Moves. The other effects, Transcription Method and Face-to-Face Effect, were not significant for Check Moves. Figure 23 summarizes the results of hypothesis testing across all five conditions[15]. Figure 24 shows the percentage of Check Moves for individual participant pairs.

[15] Visual examination of Figure 22 suggests that the interaction term of the ANOVA should be significant for CART-FTF and CART-only vs. ASR-FTF and ASR-only. However, recall that the ANOVA was run on arcsine transformed percentage scores which reduced the range of variance.

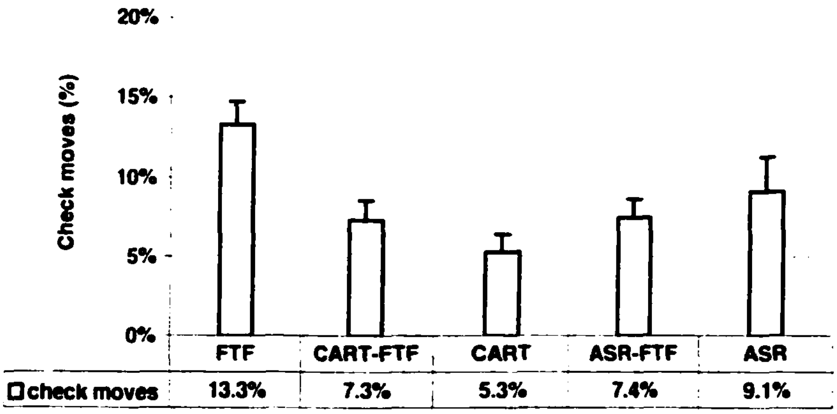


Figure 23: Summary of percent Check Moves
(mean and s.e.)

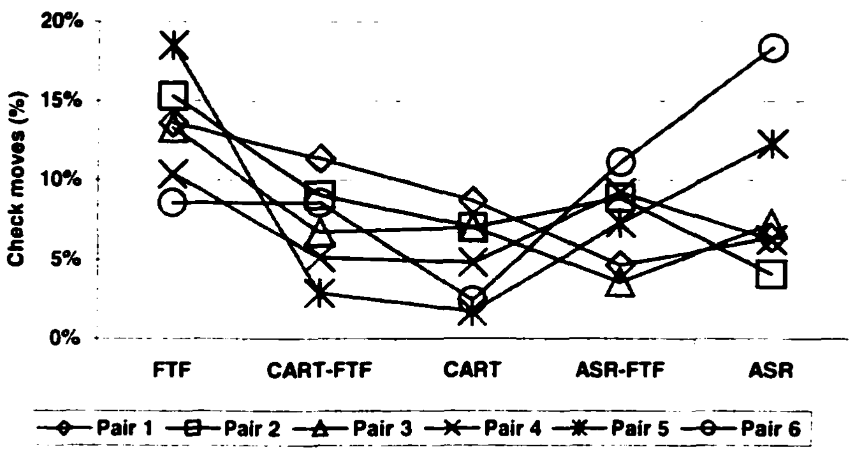


Figure 24: Percent Check Moves for individual participant pairs

Participant Evaluation Results

Upon completion of experimental sessions, participants ranked the communication conditions in terms of communicative efficiency. The ranking scale ranged from one (most efficient) to five (least efficient).

The Wilcoxon test was used to test the hypotheses. The Wilcoxon test was chosen because it tests for a difference between the medians of two related samples

(Analyse-it, 2001). It is the non-parametric alternative to the paired samples t-test. The Wilcoxon test assumes two related samples, with similar shape distributions, measured on an ordinal scale (Siegel & Castellan, 1988). The p-value is computed as follows: When the number of cases is less than 25 (as in this case), then an exact p-value is based on the assumption of no ties (Conover, 1980). If a few ties are present, the p-value only underestimates the significance by a small amount. The confidence interval around the difference between medians is computed using the Hodges-Lehman method, provided both samples are measured on a continuous scale. Median values and quartile ranges are reported for participant ranking because these statistics are more appropriate than the mean and standard deviation when reporting values measured on an ordinal scale (such as rankings). Results of hypothesis testing are discussed below.

Text-Transcription Effect

Hypothesis 1: CART-face-to-face will be ranked more favorably than face-to-face-only.

A Wilcoxon test was run to compare participant rankings in the CART-face-to-face and face-to-face-only conditions. The condition effect was significant ($p = 0.009$) (Table 24, Figure 25). Therefore hypothesis 1 was accepted.

Difference between pairs	n	Rank sum	Mean rank
Positive	10	71.0	7.10
Negative	2	7.0	3.50
Zero	0		
Difference between medians	1.500		
95.8% CI	0.500	to 2.500	(exact)
Wilcoxon's W statistic	71		
2-tailed p	0.0093 (exact tables used. 83% ties)		

Table 24: Participant rankings for Text-Transcription Effect (Wilcoxon table

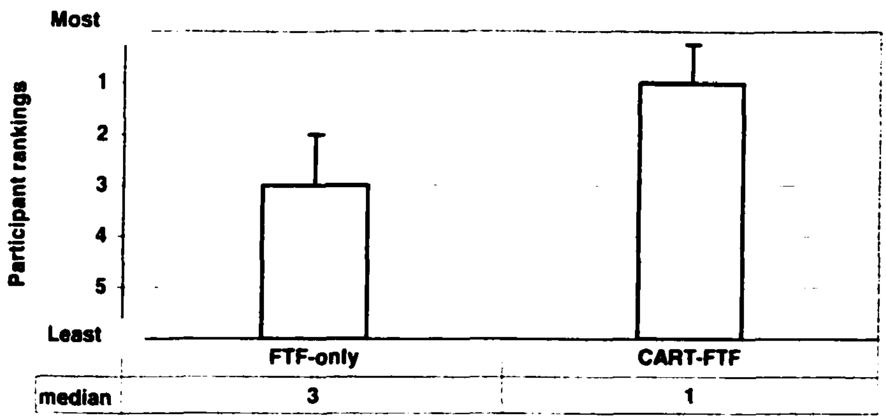


Figure 25: Participant Evaluation for Text-Transcription Effect
(median and quartile ranges)

Transcription Method effect

Hypothesis 2: CART-face-to-face and CART-only will be evaluated more favorably than ASR-face-to-face and ASR-only.

A Wilcoxon test was run to test the Transcription Method effect. The main effect was significant (p = 0.0005) (Table 25. Figure 26). Therefore hypothesis 2 was accepted.

Difference between pairs	n	Rank sum	Mean rank
Positive	0	0.0	-
Negative	23	276.0	12.00
Zero	1		
Difference between medians	-2.500		
95.1% CI	-3.000to -2.000	(exact)	
Wilcoxon's W statistic	0		
2-tailed p	<0.0001 (exact tables used. 100% ties)		

Table 25: Participant rankings for Transcription Method Effect (Wilcoxon table)

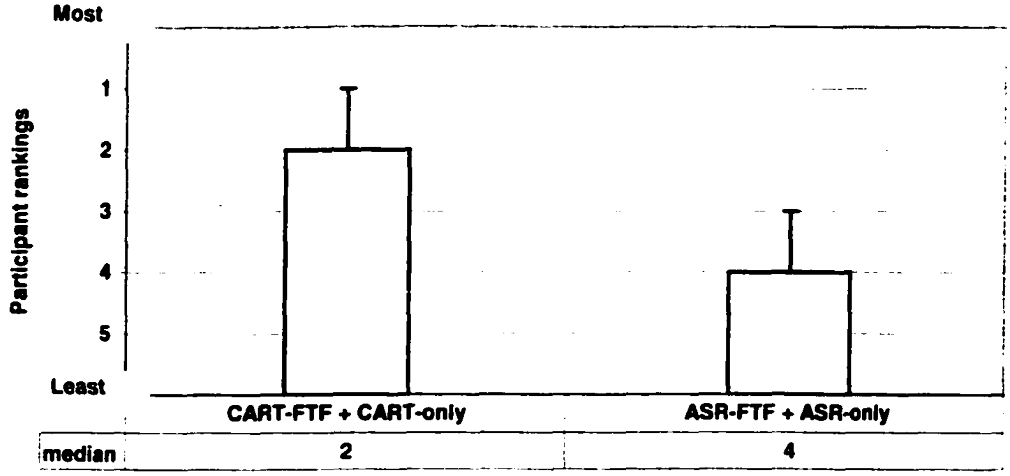


Figure 26: Participant Evaluation for Transcription Method Effect
(median and quartile ranges)

Face-to-Face Effect

Hypothesis 3: CART-face-to-face and ASR-face-to-face will be evaluated more favorably than CART-only and ASR-only.

A Wilcoxon test was run to test the Face-to-Face Effect. The main effect was significant ($p = 0.001$) (Table26, Figure 27). Therefore hypothesis 3 was accepted.

Difference between pairs	n	Rank sum	Mean rank
Positive	1	7.0	7.00
Negative	21	246.0	11.71
Zero	2		
Difference between medians	-1.500		
95.1% CI	-1.500to -1.000 (exact)		
Wilcoxon's W statistic	7		
2-tailed p	<0.0001 (exact tables used, 95% ties)		

Table 26: Participant Evaluation Wilcoxon table for Face-to-Face Effect

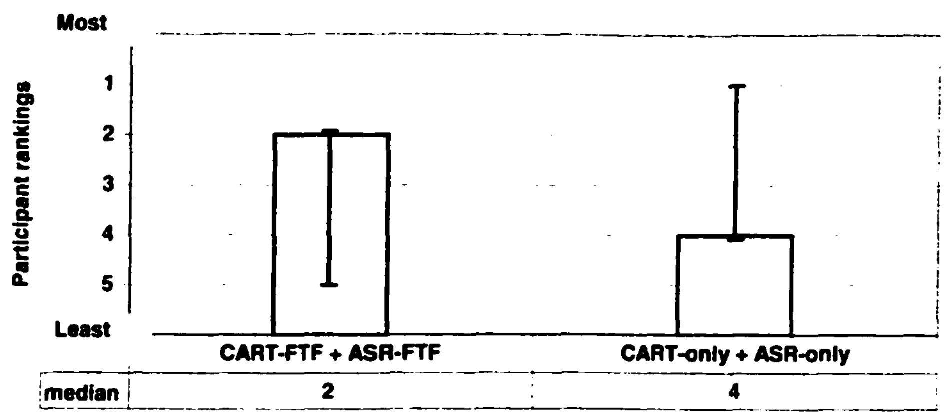
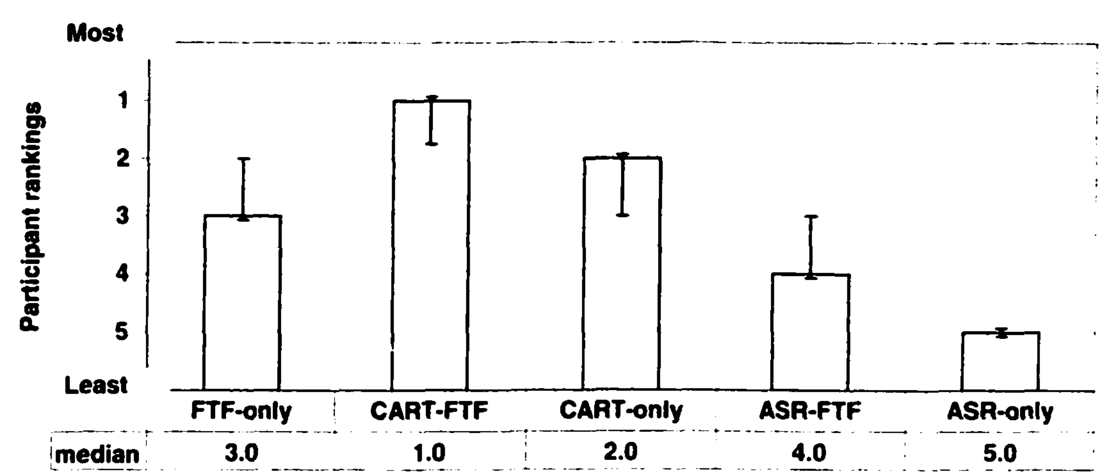


Figure 27: Participant Evaluation for Face-to-Face Effect
(median and quartile ranges)

Figure 28 summarizes the results of Participant Evaluation hypothesis testing across all five conditions. In sum, results suggest that participants preferred CART-face-to-face over face-to-face-only. The CART conditions were ranked more favorably than the ASR conditions and the conditions with face-to-face–plus-text were preferred over text-only.



ASR-only all responses were 5.

Figure 28: Participant Evaluation by condition
(median and quartile ranges)

Summary of hypothesis testing

Table 27 shows that the Participant Evaluation measure supported all three hypotheses and the other measures provided support for the hypotheses in only two cases. CART conditions were more successful than ASR conditions in terms of Completion Time and CART-face-to-face was more efficient that face-to-face-only in terms of Check Moves.

	Measures of Communicative Success			
Hypotheses	**Task Outcome**	**Time**	**Check Moves**	**Participant Evaluation**
1. CART-FTF > FTF-only	No	No	Yes	Yes
2. CART > ASR	No	Yes	No	Yes
3. Text-FTF > Text-only	No	No	No	Yes

Table 27: Summary results of hypothesis testing

It is worth noting that none of the hypotheses were supported by the Task Outcome measure and all the hypotheses were supported by the Participant Evaluation measure. It could be argued that the hypotheses do in fact hold true, as shown by the Participant Evaluation data, but that the three other measures are not sufficiently sensitive to show the effects. Another possibility is that other factors were confounding the measures of Communicative Success. Two possible factors are Transcription Accuracy and Audio-Visual Speech-Perception.

Post-hoc analyses

Post-hoc analyses looked at specific comparisons of the measures of Communicative Success that were not addressed by hypothesis testing. Post-hoc comparisons found significant differences on two of the original measures: Check Moves and Participant Evaluation. Although post-hoc analyses were conducted on the remaining measures, Task Outcome and Completion Time, the results were not significant and will not be reported here.

Post-hoc analyses also looked at two variables that may have had an impact on the results: Transcription Accuracy of the hearing participants speech and the Audio-Visual Speech-Perception abilities of the deaf participants. Transcription Accuracy is a measure of how accurately the CART and ASR speech-to-text systems transcribed the speech of each hearing participant. Audio-Visual Speech-Perception is a measure of the deaf participants' ability to perceive speech using visual cues and residual hearing. It seemed likely that these two measures would correlate with at least one measure of Communicative Success. The results of post-hoc analyses follow. These two measures were correlated with the four measures of Communicative Success.

Check Moves post-hoc

Recall that hypothesis testing revealed significant differences between percent Check Moves in the face-to-face-only and CART-face-to-face conditions ($p = 0.04$). Additional two-way repeated-measures ANOVA's were run to compare percent of Check Moves between face-to-face-only and the remaining communication conditions. Results indicated that performance in the face-to-face condition was significantly different from that of the CART-only condition ($p = 0.009$), the ASR-face-to-face condition ($p = 0.04$), but not the ASR-only condition ($p = 0.14$). See Appendix 20 for post-hoc ANOVA tables. In sum, participants used significantly more Check Moves in the face-to-face-only condition than in both CART conditions and in the ASR-face-to-face condition (Figure 29).

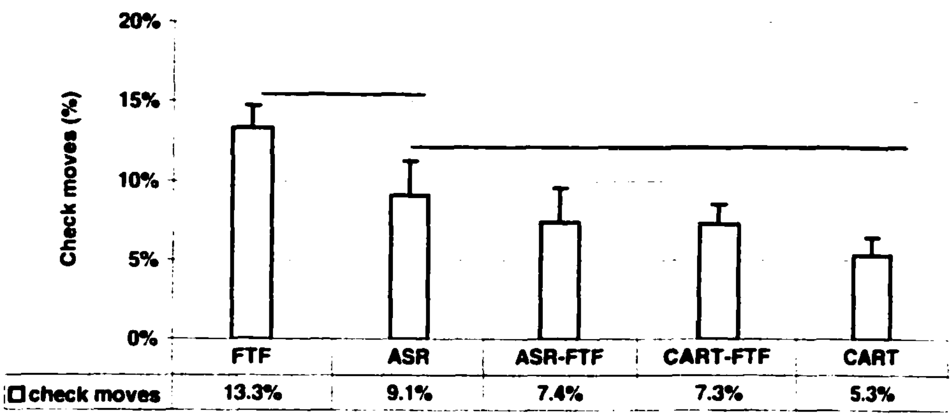


Figure 29: Post-hoc results for Check Moves
Horizontal lines indicate post-hoc groupings. (mean, s.e.)

This suggests that the text provided information which allowed participants to feel more secure that messages had been understood. The number of Check Moves in the ASR-only condition was not significantly different from that in the face-to-face-only condition, suggesting that participants were not as secure in the accuracy of the ASR transcription. Note that, in terms of Check Moves, ASR-face-to-face and CART-face-to-face conditions are indistinguishable (7.4% vs. 7.3%). A possible explanation could be that in the ASR-face-to-face condition, the ASR text was useful as a complement to the facial cues and may have provided confirmation to the deaf participant. In other words, the text served as confirmation rather than as the primary source of information.

Of the text-based conditions, the ASR-only condition had the most Check Moves and the CART-only condition had the fewest Check Moves. Although the differences were not significant, they suggested a need to further investigate performance on these conditions. Transcription Accuracy could very likely explain the difference in Check Moves when the text is the major source of information. Later in this chapter the differences in the Transcription Accuracy for the text-only conditions are discussed.

Participant Evaluation post-hoc

All hypotheses for Participant Evaluation were supported by hypothesis testing. Additional Wilcoxon tests were run to explore other specific comparisons. Results are summarized in Figure 30. (See Appendix 21 for post-hoc Wilcoxon tables.) Participant Evaluation of the CART-face-to-face condition was significantly different from CART-only condition ($p = 0.0068$). The face-to-face-only condition was not significantly different from its neighbors: the CART-only condition ($p = 0.73$) and the ASR-face-to-face condition ($p = 0.10$). The CART-only condition was significantly different from the ASR-face-to-face condition ($p = 0.02$). The ASR-only condition was significantly different from the ASR-face-to-face condition ($p = 0.001$).

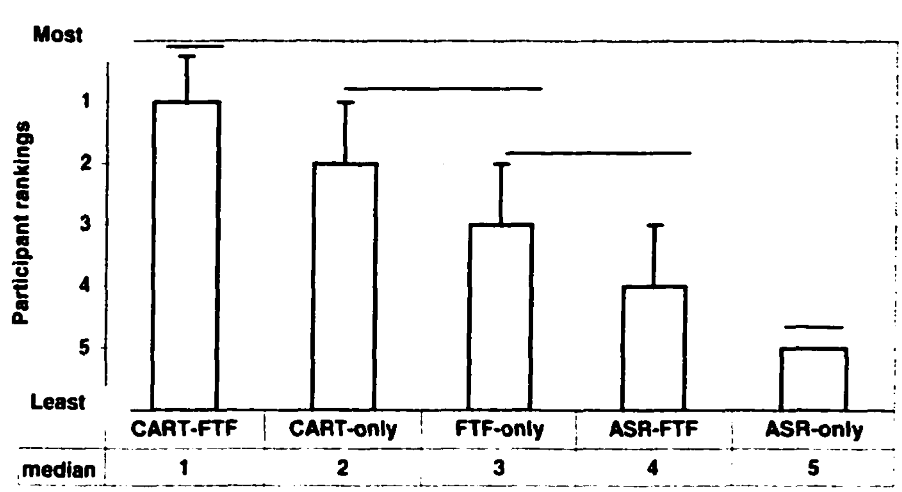


Figure 30: Post-hoc results for Participant Evaluation
Horizontal lines indicate post-hoc groupings. (median and quartile ranges)

In sum, the data fell into four groups. The CART-face-to-face condition was ranked the best and the ASR-only condition was ranked the worst. The rankings of the CART-only and face-to-face-only conditions were not significantly different from each other and the rankings of the face-to-face-only and ASR-face-to-face conditions were also not significantly different from each other.

Transcription Accuracy

Recall from the literature review that CART is a widely-used and accepted form of real-time transcription that is known to be highly accurate (Although, of course, accuracy is dependent upon the skill of the stenographer and the difficulty of the spoken material). ASR, on the other hand is a new technology that has not yet been used for real-time transcription of conversational speech. ASR accuracy is dependent, in large part, upon the speech characteristics of the talker. Transcription Accuracy was calculated based on Stuckless (1999). (See Method chapter, Transcription Accuracy section for details.)

Transcription Accuracy was calculated for the four text-based experimental conditions: CART-face-to-face, ASR-face-to-face, CART-only, ASR-only. Transcription Accuracy was also calculated for one other condition: ASR-dictation. Since the ASR software was designed for dictation-style speech, a measure of how the system performed under the conditions for which it was designed was warranted. Recall that each hearing participant had dictated a 230-word article that was transcribed by the ASR system. Mean Transcription Accuracy for ASR-dictation was 89%.

After the arcsine transformation was applied, a two-way repeated-measures ANOVA was run to investigate Transcription Accuracy differences between ASR dictation and the four experimental conditions. The main effect for condition (C) was significant ($p = 0.001$). Post-hoc testing revealed the groupings as shown in Table 28 and Figure 31.

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	SIGNIF.
Condition (C)	**1.51**	**4**	**0.37**	**15.66**	**0.001**
Participant Pairs (P)	0.22	5	0.04	1.79	0.15
CP	0.48	20	0.02		
TOTAL	2.21	29			

POST-HOC: TUKEY HSD METHOD (0.05 level) TABLE VALUE: 4.23 HSD = 0.27

SUBSET 1	GROUP	ASR-FTF	ASR-only	MEAN	2.1568	2.4046
SUBSET 2	GROUP	ASR-only	ASR-dictation	MEAN	2.4046	2.4652
SUBSET 3	GROUP	ASR-dictation	CART-FTF	MEAN	2.4652	2.6732
SUBSET 4	GROUP	CART-FTF	CART	MEAN	2.6732	2.8058

Table 28: Transcription Accuracy ANOVA table and post-hoc comparisons

In Figure 31, note the large difference between the mean values for the ASR-face-to-face and ASR-only conditions. Although it would appear that the difference between these two means should approach significance, the difference between these means is not significant because for percentages, a larger difference and smaller standard error are needed to reach significance for values around 75%, than are needed for values around 90%.

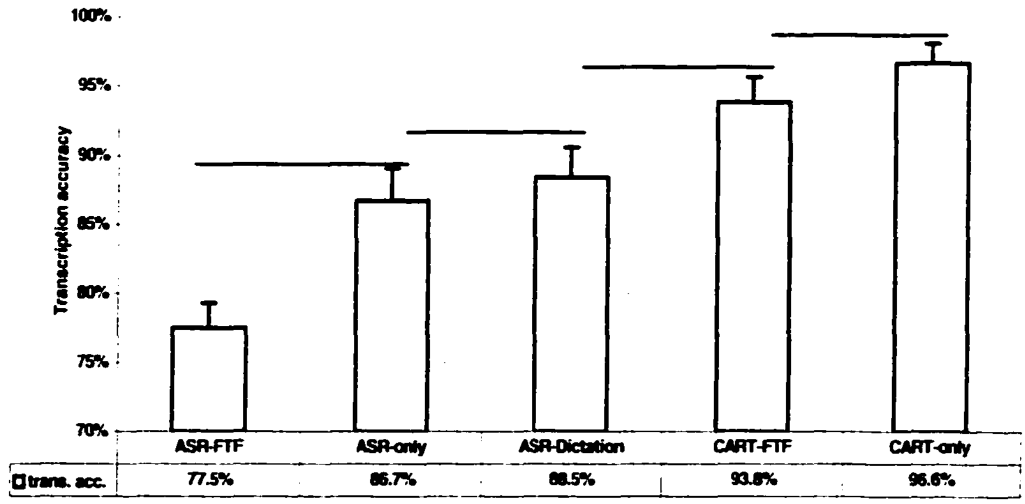


Figure 31: Post-hoc results for Transcription Accuracy
Horizontal lines indicate post-hoc groupings. (mean, s.e.)

What is interesting to note about these results is that Transcription Accuracy for the ASR-dictation condition was not as great as that for both CART conditions. In addition, CART, as a Transcription Method for conversational speech, was significantly more accurate than ASR. These differences prompt a more detailed exploration of the relationship between Transcription Accuracy and the measures of Communicative Success.

Correlations

Pearson correlations were run between Transcription Accuracy and the four measures of Communicative Success collapsed across conditions. Correlations were significant only for Participant Evaluation (Table 29).

Post hoc		Measures of Communicative Success			
Pearson Correlations		Task Outcome	Time	Check Moves	Participant Evaluation
Transcription Accuracy	All conditions	p = 0.18	p = 0.11	p = 0.10	p = 0.0006**

Table 29: Post-hoc testing of Transcription Accuracy correlations (all conditions)

Figure 32 makes it clear that Transcription Accuracy for ASR-face-to-face was poorer than that for most of the CART and ASR-only dialogues. What is interesting is that Transcription Accuracy for many of the ASR-only dialogues was the same or better than that for several CART-face-to-face and CART-only dialogues. Yet, participants ranked the ASR-only conditions unfavorably. I speculate that some other factor, like Completion Time, collaborative effort or level of confidence entered into the ranking decisions. This finding suggests that Transcription Accuracy was but one of several factors that determined how participants evaluated the conditions.

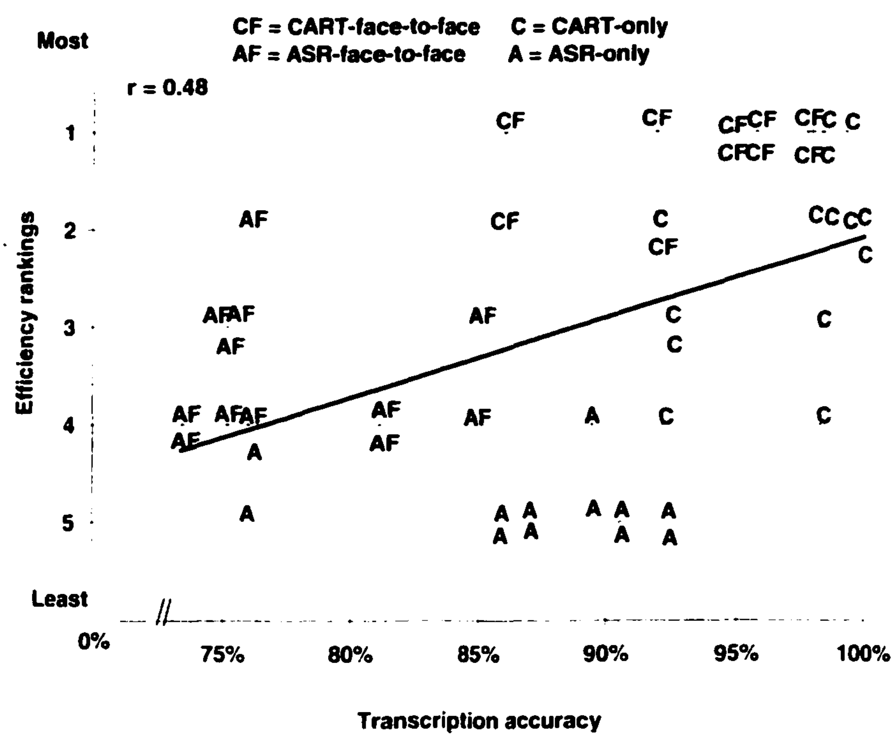


Figure 32: Transcription Accuracy and Participant Evaluation (Pearson)
(n=12 deaf and hearing participants)

Note in Figure 32 how the rankings for each group are clustered together. The Transcription Accuracy values and Participant Rankings for each condition span more or less specific ranges: ASR-only (2-4); ASR-face-to-face (4-5); CART-face-to-face (1-2); and CART-only (1-4). A Spearman analysis for each condition revealed a significant correlation for the CART-only condition ($r = 0.66$, $p = 0.02$) (Table 30). This correlation occurred because the Participant Rankings of the CART-only condition had the largest range (1-4).

	Spearman	
	rs statistic	2-tailed p
CART- FTF	0.17	0.60
CART	0.66	0.02 *
ASR- FTF	-0.02	0.94
ASR	-0.26	0.41

* Correlation significant at the $p < .05$ level.

Table 30: Transcription Accuracy and Participant Evaluation (Spearman)

I speculated that Transcription Accuracy was likely to play a greater role in Communicative Success when the text transcription was the primary source of information (i.e. in the CART-only and ASR-only conditions). Correlations were run on the text-only conditions between Transcription Accuracy and the measures of Communicative Success. Correlations were significant for Task Outcome ($p = 0.04$) and approached significance for Completion Time (($p = 0.08$) and Check Moves ($p = 0.06$) (Table 31, next page).

Post hoc		Measures of Communicative Success			
Pearson Correlations		Task Outcome	Time	Check Moves	Participant Evaluation
Transcription Accuracy	All conditions	p = 0.18	p = 0.11	p = 0.10	p = 0.0006**
	Text-only	**p = 0.04***	**p = 0.08**	**p = 0.06**	**p = 0.0002****

Table 31: Post-hoc testing of Transcription Accuracy correlations

The Task Outcome correlations in Table 31 are of particular interest. The correlations between Task Outcome and Transcription Accuracy were not significant across all communication conditions, however the correlation was significant for the text-only conditions. Figure 33 (on the next page) illustrates this point. For Task Outcome the correlation was 0.49, suggesting that more accurate transcriptions resulted in higher Map-Task scores (Figure 33). The fact that no correlation was found when the text-face-to-face conditions were included suggests that participants were getting enough information via facial cues and did not have to rely on the text (whether or not it was accurate). In the text-only conditions, the importance of Transcription Accuracy becomes clear. In general when Transcription Accuracy is greater than 90%, participants perform reasonably well. When Transcription Accuracy drops below 90%, we see that performance suffers. These results suggest that 90% is the minimum level of Transcription Accuracy needed for ASR to be a viable technology for real-time speech-to-text transcription. Further testing with larger sample sizes and alternative tasks will still be necessary to confirm this observation and determine the critical Transcription Accuracy level required for effective communication.

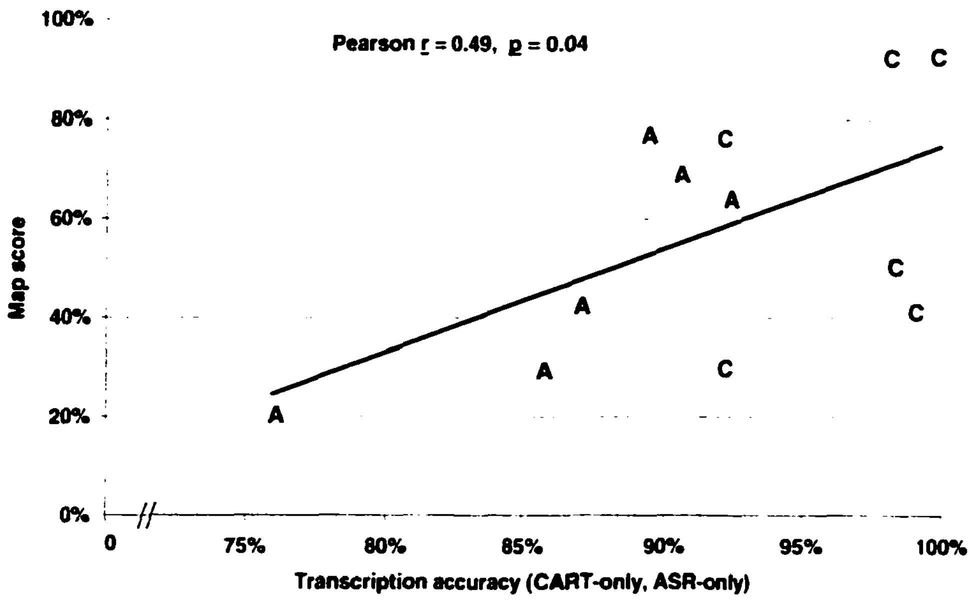


Figure 33: Transcription Accuracy and Task Outcome (Pearson)

On the next page, Figure 34, we see an inverse relationship between Transcription Accuracy and Completion Time for the text-only conditions. Correlations between the text-only conditions (ASR-only and CART-only) and Transcription Accuracy approached significance (r = -0.53, p = 0.08). Figure 34 illustrates the relationship and shows that for five out of six participant pairs in ASR-only condition, Completion Time was greater than twelve minutes. By contrast, for five out of six participant pairs in the CART-only conditions, Completion Times were twelve minutes or less. These results suggest that the poorer Transcription Accuracy in the ASR-only condition could have contributed to the longer Completion Times.

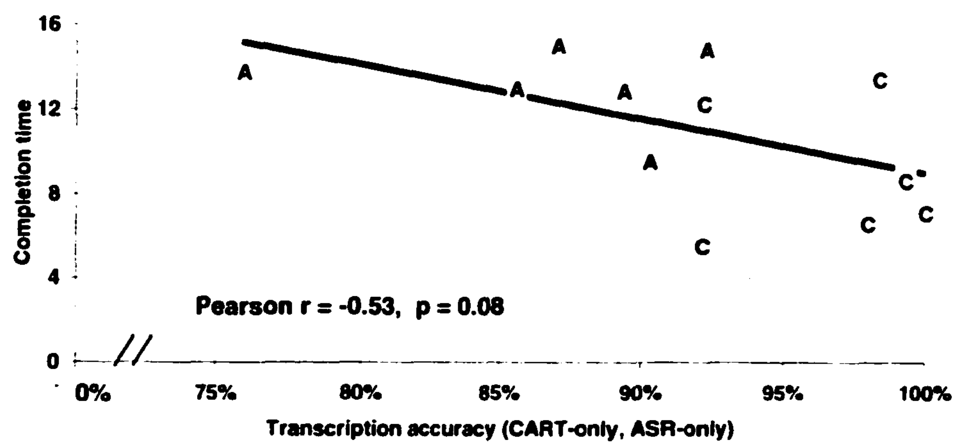


Figure 34: Transcription Accuracy and Completion Time (Pearson)

Figure 35 illustrates the relationship between CART-only and ASR-only completion times for individual participant pairs.

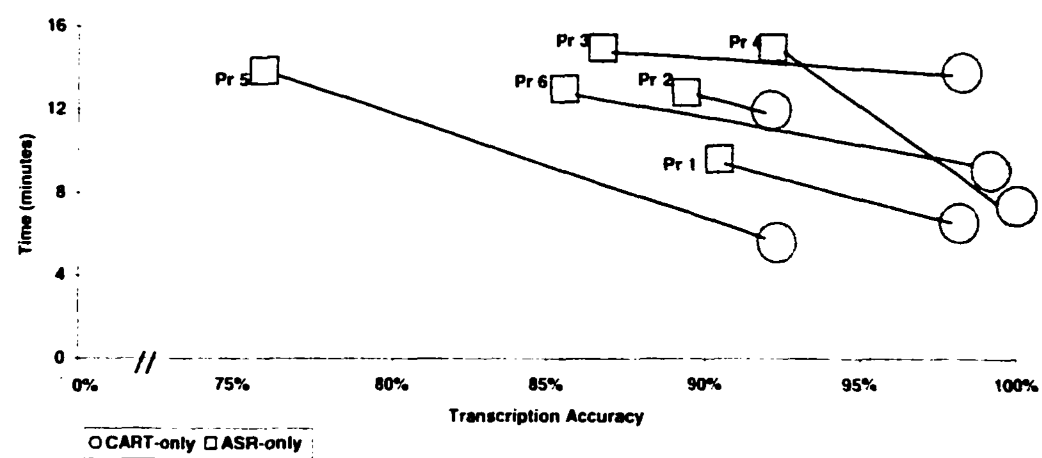


Figure 35: Transcription Accuracy and Completion Time
(Slopes from ASR-only to CART-only for individual participant pairs)

Correlations also approach significance between Transcription Accuracy and Check Moves in the text-only conditions ($r = -0.56$, $p = 0.06$) (Figure 36, next page). The correlation was negative: as Transcription Accuracy increased, Check Moves decreased.

This suggests that Transcription Accuracy did influence participant behavior in that more Check Moves were made when the transcription was less accurate.

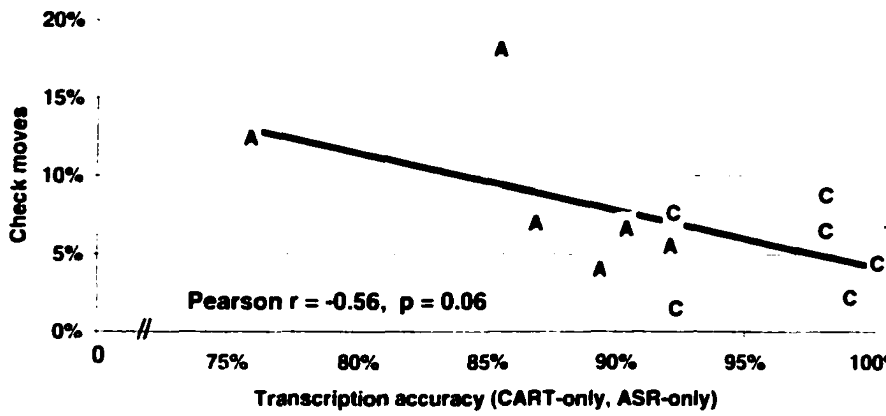


Figure 36: Transcription Accuracy and Check Moves (Pearson)

In sum, this post-hoc correlational analysis serves to further explore hypothesis 2, concerning the difference between CART and ASR conditions. The correlations for the text-only conditions show that Transcription Accuracy played an important role in Communicative Success especially when text was the major source of information. The fact that ASR tends to be less accurate negatively impacts the Communicative Success of participants using it, either alone or in conjunction with participants seeing each other.

Audio-Visual Speech-Perception and Severity of Hearing Loss

This section will address whether the deaf participants' Audio-Visual Speech-Perception ability had an effect on the experimental results. The term, *Audio-Visual Speech-Perception*, refers to understanding of spoken language using both visual cues (speech-reading) and residual hearing. This term is used instead of *speech-reading* because in this study, deaf participants were able to make use of their residual hearing in addition to speech-reading.[16] See the Methods chapter, Participants section for a detailed description of the assessment of audio-visual speech perception. The relationship between Severity of Hearing Loss and the measures of Communicative Success will be examined. Recall that four of the six deaf participants had profound hearing loss while the remaining two had severe-to-profound hearing loss.

Pearson correlations between Audio-Visual Speech-Perception and measures of Communicative Success were significant on only one measure: Task Outcome (Table 14). Spearman correlations[17] between Severity of Hearing Loss and measures of Communicative Success revealed an identical pattern (Table 32).

Post hoc	Measures of Communicative Success			
Correlations	**Task Outcome**	**Time**	**Check Moves**	**Participant Evaluation**
Audio-Visual Speech Perception	**$p = 0.0001^{**}$**	$p = 0.87$	$p = 0.66$	$p = 0.36$
Severity of Hearing Loss	**$p = 0.009^{*}$**	$p = 0.14$	$p = 0.90$	$p = 0.79$

Table 32: Post-hoc correlations for Audio-Visual Speech-Perception and Severity of Hearing Loss

These findings indicate that Audio-Visual Speech-Perception and Severity of Hearing Loss are predictors of Task Outcome on this specific task (the Map Task) regardless of communication condition (Figure 37 and 38). In other words, individual differences had a greater effect on performance than communication condition. Implications of this finding are discussed in the next chapter.

[16] Thanks to Arthur Boothroyd for advice on terminology.
[17] Spearman correlation is the non-parametric equivalent of the Pearson correlation and therefore appropriate to use with ordinal scales such severity of hearing loss.

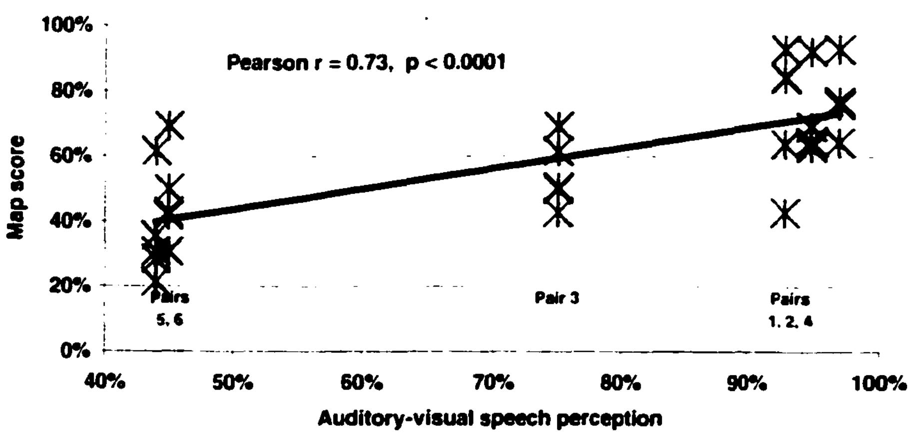


Figure 37: Task Outcome and Audio-Visual Speech-Perception

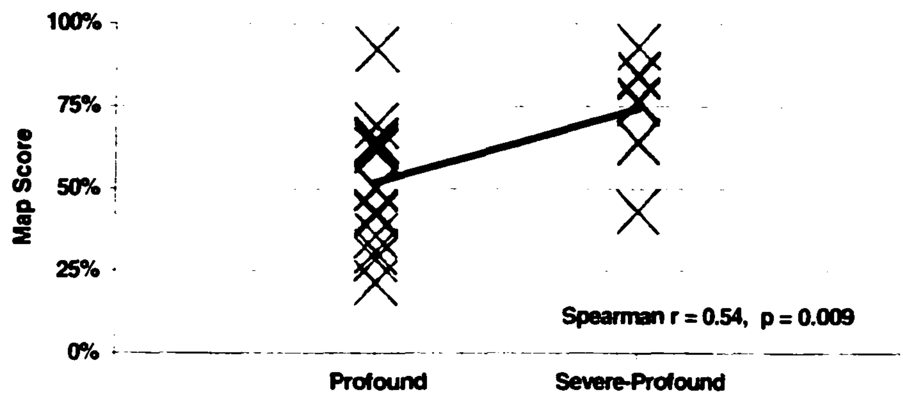


Figure 38: Task Outcome and Severity of Hearing Loss

Summary of post-hoc analyses

In sum, the post-hoc analyses give support to the notion that the factors that predict Communicative Success interrelate in complex ways. For example, the transcription accuracy of the speech-to-text technology and the deaf participants' Audio-Visual Speech-Perception abilities interact and affect Communicative Success. The discussion chapter will relate these findings back to Clark's notion of "grounding."

Discussion

The reader will recall that in the introduction the following question was raised: When two people enter into a conversation, how can the success of the interaction be measured? This question is central to evaluating the viability of technology that is *Success* was measured in the following ways:

1. Task Outcome: Did participants accomplish their goal?
2. Communicative Efficiency: How efficiently did participants communicate?
3. Participant Evaluation: How did participants evaluate the interaction?

In this chapter, I return to these issues as well as to Clark's notion of Grounding and the relationship between Grounding and Communicative Success. First, the three hypotheses are revisited and the results are discussed with respect to Grounding Theory. Then, the limitations of this experiment are discussed. Finally the implications of this research for further evaluation of technology are considered.

Hypotheses and Grounding

*Hypothesis 1. Communication will be more successful in the **CART-face-to-face** than in the **face-to-face-only** condition (Text-Transcription Effect).*

The Text-Transcription Effect was supported by Participant Evaluation and Check Move measures. In terms of Participant Evaluation, CART-face-to-face was ranked highest, i.e. best, among all conditions. In terms of Check Moves, significantly more were used in the face-to-face-only condition than in the CART-face-to face condition. Previous research (e.g. Doherty-Sneddon et al., 1997) showed that more Check Moves were associated with increased communicative effort and thus decreased efficiency. The finding that the face-to-face-only condition required more Check Moves than the CART-face-to-face condition suggests that to attain the same performance on the Map Task in the same amount of time, participants had to make more effort in the face-to-face-only condition.[18] The addition of text in the CART-face-to-face condition resulted in fewer Check Moves for an equivalent level of performance and thus greater communicative efficiency. In other words, the addition of CART text gave participants access to information that improved communicative efficiency and increased comfort in communication.

Clark's notion of **Grounding Features** may explain the differences in performance between face-to-face-only and CART-face-to-face communication. Table 33 illustrates the Grounding Features associated with these communication conditions

[18] This difference was found despite the fact that participants were permitted to type in the face-to-face condition. If participants had not been permitted to type, the effect may have been even stronger.

from the point of view of the participant with hearing loss. Only the features that are relevant to this discussion are listed in the tables.

Grounding Features	Face-to-face	CART-face-to-face
Copresence	YES	YES
Visibility (face)	YES	YES
Audibility	Partial/none	Partial/none
Cotemporality	YES	YES
Reviewability (text)	NO	Short-term

Table 33: Grounding features for people with hearing loss (ftf vs. CART-ftf)

As shown in Table 33, the face-to-face-only and CART-face-to-face conditions share the following features of 'normal' face-to-face communication: *copresence*, *visibility* (of the partner's face) and *cotemporality*. The *audibility* feature is marked 'variable' because it depends upon the individual's degree of hearing loss. The conditions differ on one feature: *reviewability* (of the text transcription). The *reviewability* feature is specific to text-based communication, including letters, faxes, and e-mail, in that the receiver of the information has the opportunity to review (re-read) the message before constructing a response. *Reviewability* is not possible with face-to-face or audio-only communication because the auditory and gestural components of the speech-signal are ephemeral and disappear immediately after they have been produced. The fact that CART-face-to-face communication includes the *reviewability* feature may explain why fewer Check Moves were employed in the current study and also why both deaf and hearing participants preferred this communication method. Moreover, the CART transcription was highly accurate and reliable. For the deaf participant, not only is it preferable that the utterances of the hearing partner are *reviewable*, but the text must also be accurate.

*Hypothesis 2: Communication will be more successful in the **CART-face-to-face** and the **CART-only** condition than in the **ASR-face-to-face** and the **ASR-only** condition (Transcription Method Effect).*

The Transcription Method Effect was supported by the Completion Time and Participant Evaluation measures. In terms of Participant Evaluation, both CART conditions were ranked significantly more favorably than both ASR conditions. With respect to Completion Time, the ASR conditions required significantly more time than the CART conditions. This effect was significant despite the fact that a 15-minute time limit was imposed. Post-hoc analyses of Transcription Accuracy provide a likely explanation. The ASR transcriptions had significantly more errors than the CART transcriptions. It is not surprising, therefore, that participants preferred CART. Moreover, the error-laden ASR conditions likely required more time because hearing participants spent time and energy attempting to correct ASR errors as they appeared. Informal observations during data collection suggested that when normal hearing participants did attempt to correct, they became frustrated when the system responded unpredictably to correction attempts. Two examples of ASR errors and repairs follow.

Example 1: ASR error and repair (Pair 4 A5)

Hearing Participant: Is <u>there</u> anything <u>near</u> the bridge besides the banana tree <u>and</u> the camp?

ASR: Is anything <u>in</u> the bridge besides the banana tree <u>in</u> the camp

Deaf Participant: There's nothing in the bridge.

Hearing Participant (typed): near

Deaf Participant: Oh, near the bridge.

Example 1 is an instant when the hearing participant successfully resorted to typing to correct an ASR transcription error: "in" for "near."

Example 2: ASR error and repair (Pair 1 AF4)

Hearing Participant:	The line moves into the right.
ASR:	bowline moves into the right
Deaf Participant:	Bowline moves to [r]. I don't know what a bowline is.
Hearing Participant:	The line goes.
ASR:	bowline goes
Deaf Participant:	The line. Bowline. The line.

In Example 2, the ASR system consistently transcribed "the line" as "bowline." The hearing participant attempted to correct the error by repeating the utterance, but to no avail. However, the deaf participant then gave feedback indicating his/her understanding that "bowline" meant "the line." Both of these examples illustrate how participants adapted to the constraints of ASR and worked around ASR transcription errors. This is important because it shows that increased communicative effort was necessary to insure Communicative Success when using ASR.

Post-hoc analyses correlating Transcription Accuracy and Communicative Success for the text-only conditions (CART-only and ASR-only) are illustrative here. Correlations were significant or approached significance across all measures of Communicative Success, suggesting that when facial cues were not available the differences between CART and ASR were more pronounced.

An important future contribution would be to calculate a minimum ASR accuracy level that is viable for real-time transcription of conversational speech. An eyeballing of the correlations between Transcription Accuracy and Participant Evaluation in the text-only conditions suggests that, on tasks like the one employed here, transcription must be better than 90% in order for participants to favorably judge the communicative interactions. These results suggest that 90% is the minimum level of Transcription Accuracy needed for ASR to be a viable technology for real-time speech-to-text transcription. This claim must, of course, be verified by further inquiry.

The state-of-the-art of ASR technology is simply not forgiving enough of the normal range of human speech errors, such as dysfluencies and hesitations. The CART stenographer, on the other hand, can hear a speech error and simply ignore it. Because of this limitation, ASR systems at this time require that the talker use 'careful speech'. In this study, Transcription Accuracy was much poorer in the ASR-face-to-face condition. Hearing participants tended to speak slowly, clearly and with aforethought in the text-only conditions and ASR-dictation conditions. The ASR-face-to-face condition was more difficult for the hearing participants because of divided attention. They had to use their eyes to monitor the ASR output and as well as make eye contact for communication with their deaf partners. Personal observation of the data collection led me to believe that the reason for the poorer transcription accuracy in the ASR-face-to-face condition was because hearing participants tended to speak more "naturally." In other words, they tended to produce more hesitations and dysfluencies in the ASR-face-to-face condition than the ASR-only condition.

A few examples follow, illustrating the types of errors the ASR system made when transcribing hesitations and dysfluencies.

Example 1: ASR dysfluency transcription error (Pair 5 AF4)

Hearing Participant: [lə] Let me describe to you

ASR: will let me describe to you

Example 2: ASR dysfluency transcription error (Pair 2 A3)

Hearing Participant: [N] No. Yeah, yeah yeah, there's a railroad crossing.

ASR: and now that he out the other are across

*Hypothesis 3: Communication will be more successful in the **CART-face-to-face** and **ASR-face-to-face** conditions than in the **CART-only** and **ASR-only** conditions (Face-to-Face Effect).*

The Face-to-Face Effect was supported by the Participant Evaluation measure. Participants ranked the face-to-face conditions significantly more favorably than the text-only conditions. These differences were not supported by the other measures of Communicative Success. One possible reason is that the differences on other measures between text-plus-face-to-face and text-only were too small to reveal significant differences with a sample size of six pairs. A second reason is that the differences between the Transcription Methods (CART and ASR) may have been so large as to mask subtler differences between face-to-face-plus-text and text-only.

However, a difference between text-face-to-face and text-only was evident in the post-hoc analyses. Correlations between Transcription Accuracy and the three remaining measures of Communicative Success (Task Outcome, Completion Time, and Check Moves) were significant for the text-only conditions. When the text-face-to-face conditions were added to the analysis, the correlations were no longer significant. This suggests that when facial cues were available, participants tended to rely on them more than the text. The text was used as a backup, but not the primary source of information transfer.

Clark's grounding features may again be instructive to explain the differences between text-face-to-face and text-only communication. Table 34 illustrates the grounding features associated with these communication conditions from the point of view of the participant with a hearing loss.

Grounding features	Text-face-to-face	Text-only
Copresence	YES	YES
Visibility (face)	YES	NO
Audibility	Partial/none	Partial/none
Cotemporality	Variable	Variable
Reviewability (text)	YES	YES

Table 34: Grounding features for people with hearing loss (text-ftf vs. text-only)

Text-only communication differs from text-face-to-face communication by one feature, *visibility*. This feature refers to *visibility* of the face, not the text. Participants tended to prefer to see their partner's face and the text, rather than just the text. However the presence or absence of this feature did not influence Task Outcome or Communicative Efficiency. At this point we only know that participants prefer to see their partner's face, but discourse analysis measures did not reveal differences between text-only and text-face-to-face conditions. Other measures (e.g. back-channels and overlaps) may reveal differences between conditions.

In addition, the feature *cotemporality* has been marked as variable. Recall that for *cotemporality* to be present, the message must be received essentially at the same time that it is produced. For example, spoken communication is cotemporal, but faxes are not. For both ASR and CART, cotemporality will depend, to some degree, up the speaking style of the talker. For example, rate of speaking, word choice and dialect differences of the talker will affect how fast the speech can be transcribed. In the case of CART transcription, degree of cotemporality depends upon the speed and skill of the stenographer. For ASR, cotemporality depends upon the speech-to-text algorithms, the frequency with which speakers make corrections. The ASR transcription algorithms hold speech in a buffer until a complete sentence has been recognized. From the vantage point of the reader of the text (the deaf participant) sentences appear on the monitor all at once, not one or two words at a time. If the talker (the normal hearing participant) speaks too quickly or too slowly or pauses in unexpected places the transcription algorithm may be even more delayed. This variability in *cotemporality* may be another reason that participants preferred CART to ASR. Investigation of the time delay between speech and text merits further investigation.

Another issue affecting *cotemporality* and *reviewability* is the difference in ASR Transcription Accuracy between ASR-face-to-face (78%) and ASR-only (87%). Although the difference was not statistically significant ($p > .05$), it is worth discussing here. Based on personal observation, it seemed that in the ASR-only condition, the hearing participants tended to speak more carefully and monitor the text output more vigilantly than in the ASR-face-to-face condition. In the ASR-face-to-face condition, normal hearing participants had to divide their attention between monitoring their speech-to-text output and making eye contact with their partner. It appears that the strain of dividing attention between these two tasks resulted in less careful speech and a reduction in Transcription Accuracy. The effect of divided attention and ASR merits further investigation.

Limitations/Directions for future research

Although this study answered some basic questions about the viability of ASR for communication enhancement, it left many questions unanswered. Limitations of this study and areas that are likely to be fruitful for future research follow.

Sample size

A larger sample size should serve to confirm some of the significant results. It is also possible that the differences that approached significance, such as Task Outcome, Check Moves, Time to Completion and Transcription Accuracy, would become significant.

Between-participant variability

Severity of Hearing Loss and Audio-Visual Speech-Perception ability varied greatly among the deaf participants. The post-hoc analyses showed that, indeed, Task Outcome correlated with both Severity of Hearing Loss and Audio-Visual Speech-Perception ability suggesting that the range in performance on the Map Task was influenced by individual differences on these measures. A more uniform sample might be preferable for further study. For example, a group of deaf participants with more similar levels of Hearing Loss and Audio-Visual Speech-Perception ability may provide more insight into the efficacy of the speech-to-text systems for that specific sub-group. The deaf and hard-of-hearing populations have many sub-groups whose needs must be evaluated as separate groups. One interesting prediction concerns the effect of Audio-Visual Speech-Perception ability. The results of this study suggest that individual participant pairs tended to perform the same on the Map Task no matter what the condition. Individual differences across pairs were greater than within-subject differences. Originally I had imagined that poor speech-readers would show the greatest benefit and need for speech-to-text technologies. However results suggested that performance on the Map Task did not change significantly across conditions. It would be interesting to collect data using another task and test only poor speech-readers. In order to understand the effects of a text-supplement for poor speech-readers, it would be preferable to test only participants who fall into this category.

Another variable in participant-selection was age of onset of deafness. All deaf participants produced speech that was understandable for face-to-face communication. Four were late-deafened with onset of hearing loss after age 21. Two were early-deafened. One participant (3D) was prelingually-deaf and another (1D) began to lose her hearing at age 3. A comparison of the data based on age of onset of deafness did not reveal any clear-cut performance differences

Four of the six deaf participants reported profound hearing loss while two reported severe hearing loss. In this regard, a drawback of this study was the design of the text-only conditions. Since a barrier was simply placed between the deaf and hearing participants, it is likely that the participants with severe hearing loss were still able to hear some of their partner's speech. An improved design, one that would truly measure Communicative Success in a text-only condition, would require the participants to be placed in different rooms and provide them with technology to communicate remotely.

Task

The Map Task may not have been sensitive enough to show differences in performance because there may have been too much variability across maps. It is also possible that the scoring system was not sensitive enough. A task more similar to an actual workplace task may prove more instructive. For example, in a video-mediated communication study, Newlands et al. (1996) used a scenario in which one participant acted as 'travel agent' and the other participant was a customer making travel plans. Another task may provide different information about Communicative Success.

Completion Time

Another limitation related to the task was the Completion Time measure. Three of the 30 Map Tasks were not completed in the allotted fifteen minutes. To get a more accurate measure of Completion Time, it might have been preferable to allow participants as much time as they needed. We chose not to because of our own time constraints, participant fatigue and cost of the CART reporter. Although, Completion Time and the other measures of Communicative Success could have extrapolated based on how much of the maps participants had completed, this was not done due to high inter-pair and intercondition variability. In addition, one cannot assume that participants' performance would be uniform across from the beginning to the end of a given Map Task.

Discourse analysis

A more in-depth discourse analysis of the existing data may reveal interesting patterns. As described in the literature review above, studies by Doherty-Sneddon et al. (1997) and Newlands et al. (1996), showed differences among communication media on discourse measures such as interruptions and back-channels. My guess is that an analysis of the dialogues from this dissertation will reveal more interruptions and back-channels in the face-to-face conditions.

An in-depth analysis of the Check Moves may also be interesting. Check Moves are similar to Repairs and Requests for Clarification as discussed by Tye-Murray et al. (1994) and Caissie & Rockwell (1993). When those researchers subdivided repairs into several sub-types, they found that specific requests for clarification were likely to be more successful at repairing communication breakdowns than non-specific requests. An analogous in-depth analysis has not yet been applied to Check Moves in the literature and will likely prove interesting.

An analysis of eye gaze may also give information about the issue of divided attention by describing when the deaf participants looked at the text and looked at their hearing partners and when the hearing participants monitored the ASR output and maintained eye contact with their deaf partner.

Versions of ASR software

Data collection for this study was completed in December, 1999. Since then, improved versions of ASR software have been released. One version claims to have an algorithm that rejects hesitation and filler utterances such as 'um' and 'uh' (Baig, 2000). Newer versions may make more allowances for the variability and inaccuracies of conversational speech. Future work could evaluate the newer ASR versions to see if the Transcription Accuracy is improved in a conversational setting.

The following points were suggested by Harkins (in press and personal communication).

Training

The author has found no published studies examining the effects of length and quality of ASR training. A study is needed to determine if more extensive training of the ASR system would improve ASR usability and increase Transcription Accuracy. Training should include how to have a conversation using 'careful speech', how to correct errors, when to resort to typing.

Another possibility is to evaluate systems based on differences in expertise/experience of the hearing partner, not only in ASR, but also in communicating with deaf individuals. An idea for a longitudinal study would be to see how individual deaf and hearing pairs develop communication strategies over "practice " communicating with each other over several months or a year.

Keyboard skills

Only participant pairs where the deaf participant was a poor speech-reader used the keyboard. This is an area for further study: the use of the keyboard during conversation. Hearing participants' typing skills may be a factor to explore in more detail. A study could measure the effects of training people to skillfully use the keyboard as a supplement to speech-reading.

Line of sight

A study is needed to determine the optimal placement of the monitor for the text output and the optimal font size. There has been some talk of eyeglasses with mounted display. The viability of this technology could also be evaluated, because although it has been established that text does benefit communication for people with hearing loss, further study may suggest optimum placement for the text and optimum font size.

Automated manual transcription

Another idea for an ASR application is for one highly trained speaker to take the place of a stenographer or typist. Instead of typing the proceedings of a meeting or conversation, this person would speak what others had spoken into the ASR system. This method has been examined by Stuckless (2000) and is in trial stage for use by the telephone relay service (TRS) (Harkins & Bakke, in press).

Conclusions

An important contribution of this work is the application of an innovative approach to evaluate Communicative Success with and without text as a supplement to face-to-face communication. The Map Task required interlocutors to work together toward a common goal. Since the task was very specific, it elicited a manageable range of utterance types. An open-ended conversation, on the other hand, could yield a near-infinite number of utterance types and make comparative analysis next to impossible. The Map Task or a similar technique could be applied to the assessment of communication aids in general (including hearing aids) and provide new ways of assessing the performance of assistive technologies in the real world.

This project has encouraged me to develop my interest in the more general problem of developing practical outcome measures. I have come to believe more firmly in the superiority of a multi-faceted approach toward the assessment of Communicative Success. By this, I mean that the results of several measures of assessment must be taken together to give an overall picture of communication using a given device. In this study, the measures reported were Task Outcome, Completion Time, Check Moves and Participant Evaluation. In the future, other measures could be assessed such as conversational style. As mentioned in the literature review, Tye-Murray et al. (1995) proposed a model that identified three overarching 'types' of conversational styles of persons with hearing loss: interactive, passive and aggressive. According to Tye-Murray, the preferred style, interactive, is characterized by equal share of the conversational floor, low occurrence of communication breakdowns and the appropriate use of repair strategies. While I agree that individuals have personal characteristics that they bring to a conversation, individuals' conversational styles change depending on the situation and their conversation partner. This view is consistent with Grounding Theory, which asserts that conversation is an ongoing collaboration between two individuals who bear mutual responsibility for its outcome. A baseline conversational style of a particular pair of individuals could be assessed in face-to-face communication. Then one could report how the style changed when technological aids were added. Of course, the measurement of such variables as control of conversational floor and frequency of occurrence of communication breakdowns needs to be clearly thought through and defined.

In sum, this study supported the notion that for people with hearing loss, the availability of accurate real-time transcription of conversational speech improved Communicative Success. CART transcription was accurate and was ranked favorably by participants. Accuracy of ASR transcription will need to be improved before it is comparable to CART as a viable as a method of real-time speech-to-text transcription.

APPENDICES

Appendix 1: Informed Consent Form

My name is Anita Haravon. I am a doctoral student at the City University of New York in the Speech and Hearing Sciences Program. For my doctoral thesis I am investigating the potential use of automatic speech recognition (ASR) technology as communication aid for people with hearing loss. I am asking you to participate in this study in which people with hearing loss and people with normal hearing will communicate using the ASR technology.

If you have a hearing loss, you will complete a brief questionnaire about the types of assistive technology you use on a regular basis and a test to assess how much benefit you tend to receive from speech reading.

If you have normal hearing, you will train with the ASR software so it can better understand your speech.
Then, you and another participant will complete the Map Task, a task in which you tell each other about a route on a map in different ways. For example, for one round you will be able to see each other and for another round you will only see the ASR computer screen. The entire study should take approximately 4 hours. There will be breaks between sessions.

The experimental sessions will be video and audio taped for analysis purposes. Recordings will not be used for presentations without your signature on a separate release form. You will have the opportunity to review the completed recordings and ask that they not be used. Tapes will be stored in a locked cabinet at Gallaudet University and will be destroyed at the end of the project or after 3 years, whichever comes first.

You will be paid $8/hour. If you do not complete the project you will be paid a minimum of $8. Taking part in this study is voluntary. You may withdraw at any time. Your participation or withdrawal from this study will in no way affect your access to services at Gallaudet University, if you are at Gallaudet, or your employment status at Lexington, if your are at Lexington. Your responses and any personally identifiable material will be kept confidential. All participants will be assigned a number for the purposes of recording and reporting the data.

If you have any questions regarding this study, please contact me at the Lexington School for the Deaf 718-899-8800 (voice/TTY), anita.haravon@hearingresearch.org or my advisors: Harry Levitt, Ph.D. at the CUNY Graduate Center: 212-642-2359 (voice), harry.levitt@hearingresearch.org or Judy Harkins, Ph.D. at Gallaudet University: 202-651-5677 (voice/TTY), judy.harkins@hearingresearch.org.

If you have any questions about your rights as a participant in this study, you may contact Hilry Fisher, Sponsored Research, Graduate School/City University of New York: 212-642-2059 (voice). TTY callers please use NY Relay Service: 800-662-1200.

If you agree to participate, please fill out the requested information below.

Print Name: ____________________________ Date of Birth: ____________________

Phone: _______________________________ Email:___________________________

Participant's Signature: ______________________________ Date: _______________

Principal Investigator: _______________________________Date: _______________

Appendix 2: Technology Use Questionnaire: Participants with Hearing Loss

We are seeking adult-deafened people who rely primarily on visual methods of receiving communication. to participate in this study. If you fit this description and are interested in participating, please complete this questionnaire as soon as possible and return it along with a copy of your most recent audiogram (if available) in the enclosed envelope. If there are any questions that do not apply to you, please write NA (not applicable) in the blank.

Name: ______________________________ Today's Date: ____________

Age: ___________________________ Gender: Female _______ Male _____

Address: ____

Daytime Phone #: ____________________________Circle One: Voice TTY

Evening Phone #: ____________________________Circle One: Voice TTY

Fax #: ______________________ E-mail Address: __________________

Circle the best way to reach you: Daytime phone Evening phone E-mail

---

1. What is the level of hearing loss in your right and left ears? (check one for each ear)

Left Ear
_______ Mild (30 dB or less)
_______ Moderate (31-50 dB)
_______ Moderately severe (51-70 dB)
_______ Severe (71-90 dB)
_______ Profound (90 dB or more)
_______ Don't know

Right Ear
_______ Mild (30 dB or less)
_______ Moderate (31-50 dB)
_______ Moderately severe (51-70 dB)
_______ Severe (71-90 dB)
_______ Profound (90 dB or more)
_______ Don't know

2. If you are unable to send a copy of your most recent audiogram now, would you be able to bring your audiogram with you if you are selected as a participant?__ _______Yes _______No

3. How long have you been aware of your hearing loss?____ years If less than one year:_____months)

4. Do you wear hearing aids?

Left Ear:_______ Yes _______No If yes, for how long? ____ years/ _______months
Right Ear: _______ Yes _______No If yes, for how long? ____ years/ _______months

5. Do you have a cochlear implant?

Left Ear:_______ Yes _______No If yes, for how long? ____ years/ _______months
Right Ear: _______ Yes _______No If yes, for how long? ____ years/ _______months

6. Do you use the voice telephone or TTY for most of your phone calls (check the one that describes how you use the phone on most calls):

____ Voice	____ TTY
____ without a hearing aid	____ Two-way TTY (text to text)
____ with hearing aid on M (microphone)	____ Voice carry over (CVO)
____ with hearing aid on T (t-coil)	

7. When you attend a meeting, what method(s) do you use for understanding what is said (please check all that apply):

_____ Hearing aids/Cochlear implant	_______ Assistive Listening System
_____ Lipreading	_______ Sign language interpreter
_____ Real-time transcription service (CART)	_______ Oral interpreter
_____ Typist taking notes on computer (CAN)	

_____ Other (please specify:) ____

8. When you are having a face-to-face, one-on-one conversation with a hearing person:

   a. Do you usually wear a hearing aid and/or assistive listening device? ___Yes ___No

   b. Do you and your conversational partner usually use sign language? ___Yes ___No

   c. Do you usually find it easy to understand what the other person is saying?

   ___ Very easy ___ Easy ___ Somewhat easy ___ Difficult ___ Very difficult

   d. How much do you usually rely on lipreading and listening?

   ___ I rely completely on lipreading. Listening does not help at all.
   ___ I rely mostly on lipreading. Listening only helps somewhat.
   ___ I rely on lipreading and listening equally.
   ___ I rely mostly on listening. Lipreading only helps somewhat.
   ___ I rely completely on listening. Lipreading does not help at all.

   e. Are you generally satisfied with communications that are one-on-one? ___Yes ___No

9. How familiar are you with Automatic Speech Recognition (ASR)? (circle one)

   Extremely familiar   Very familiar   Familiar   Somewhat Familiar   Not Familiar

10. Have you ever used an Automatic Speech Recognition system before? ___ Yes ___ No

If yes, please describe your experience: ____

____

11. What percentage of spoken words do you think Automatic Speech Recognition systems can transcribe accurately?

___ Less than 50%   ___ 50-80%   ___ 81-90%   ___ 91-95%   ___95-100%

THANK YOU!

We will contact you to let you know whether you are selected for the study.

Appendix 3: Dictation Article

DeLorenzo, K. (March 3, 1999). New President's Fellowship Program to address need for deaf faculty in higher education, *On the Green*, p. 4, Gallaudet University.

As part of an ongoing effort by Gallaudet to increase the number of qualified deaf professors, the University has begun the process of implementing the Gallaudet University President's Fellowship Program, slated to begin this fall. Announced during the 1998 Deaf President Now Tenth Anniversary celebration, the Fellows Program is intended to recruit and support deaf and hard of hearing graduate students pursuing full-time study toward doctorate or terminal degrees who aspire to a teaching and research career. Under the auspices of the Fellows Selection and Review Committee, five fellows each year receive tuition support of up to $12,000, an annual stipend between $14,000 and $24,000, and additional funding support for research, library privileges, and travel support. Each fellow serves as a teaching assistant in his or her academic discipline, and teaches up to two courses per year.

Serving on the selection committee are representatives from each college of the University, a representative from faculty governance, one dean, and a representative from the Graduate School. The committee will begin reviewing applications in mid-April and make final selections by June 1. Regular employees of the University are not eligible, and future employment of fellows at the University is not guaranteed.

Satisfactory academic progress toward the degree and teaching performance are requisites for reappointment. Each fellow participates in New Faculty Orientation the first semester of appointment, and attends faculty development mentoring activities.

Appendix 4: Landmark names changes

Map 0

Landmark	Scotland Term	USA Term
1	camera shop	
2	parked van	parked truck
3	yacht club	
4	allotments	farm land
5	flight museum	airplane museum
6	west lake	
7	disused monastery	monastery ruins
8	alpine garden	herb garden
9	youth hostel	
10	level crossing	railroad crossing
11	telephone box	telephone booth
12	thatched mud hut	thatched roof hut
13	east lake	
14	picker fence	picket fence
15	parked van	parked truck

Map 1

Landmark	Scotland Term	USA Term
1	crest falls	waterfalls
2	banana tree	
3	footbridge	
4	poisoned stream	polluted stream
5	old temple	
6	slate mountain	stone mountain
7	abandoned truck	
8	baboons	
9	white mountain	
10	lemon grove	
11	pyramid	
12	cobbled street	cobble stone street
13	remote village	isolated village
14	poisoned stream	polluted stream

Map 2

Landmark	Scotland Term	USA Term
1	diamond mine	
2	wagon wheel	
3	rift valley	death valley
4	cactus	
5	swamp	
6	rocks	
7	stone creek	
8	white water/rapids	
9	manned fort	pioneer fort
10	outlaws hideout	
11	gold mine	
12	stone slabs	flat rocks
13	noose	
14	swan pond	
15	stone creek	
16	saloon bar	saloon

Map 3

Landmark	Scotland Term	USA Term
1	green bay	
2	sandstone cliffs	
3	forge	blacksmith
4	old pine	
5	pine forest	
6	forest stream	
7	wheatfields	wheat fields
8	bakery	
9	canal	
10	crane bay	south bay
11	wheatfields	wheat fields
12	rocket warehouse	warehouse
13	cave	
14	old lighthouse	

Map 4

Landmark	Scotland Term	USA Term
1	site of forest fire	forest fire site
2	picnic site	
3	adventure playground	
4	farmyard	
5	privately owned fields	
6	granite quarry	
7	train crossing	railroad crossing
8	waterfall	
9	public footpath	hiking trail
10	fallen cairn	fallen rocks
11	limestone cliffs	
12	lion country	
13	east lake	
14	corn fields	
15	train crossing	railroad crossing

Map 5

Landmark	Scotland Term	USA Term
1	missionary camp	
2	gorillas	
3	banana tree	
4	rope bridge	
5	straight river	
6	waterfall	
7	lost steps	hidden steps
8	fallen pillars	
9	flamingoes	
10	ancient ruins/ruined city	
11	stones	
12	white mountain	
13	soft furnishings store	furniture store
14	rock fall	fallen rocks
15	lost steps	hidden steps
16	gazelles	

Total number of original Scottish terms for all maps: 90 Total number of altered words: 34

Percent of words that have been changed: 37.8%

Appendix 5: Map Task Instructions for Deaf Participants

The map in front of you has landmarks, but no path. Your partner has a map with landmarks and a path. Without looking at each other's maps, you and you partner must work together so that you can draw the path as it appears on your partner's map.

Some of the landmarks on your partner's map may not be exactly the same as yours. They may have different names or be in different places on the map. The 'START' position on both of your maps is always the same.

You can use any means necessary to accomplish this task: ask questions, describe your map, whatever strategy seems to work best for you.

Please don't hesitate to ask your partner for clarification, if anything is unclear. It is important to draw the path as accurately as possible.

You and your partner will work on a map for each of the communication conditions. We will stop each session after 15 minutes. It's OK if you don't make it to the finish line. The first session will be a practice session.

Thank you for your cooperation.

Good Luck!

Appendix 6: Map Task Instructions for Hearing_Participants

The map in front of you has landmarks and a path. Your partner has a map with landmarks, but no path. Without looking at each other's maps, you and you partner must work together so that your partner can draw the path on his or her map.

Some of the landmarks on your partner's map may not be exactly the same as yours. They may have different names or be in different places on the map. The 'START' position on both of your maps is always the same. The 'END' position is not marked on your partner's map.

You can use any means necessary to accomplish this task: give directions, ask questions, describe your map, whatever strategy seems to work best for you.

You and your partner will complete a map for each of the communication conditions. We will stop each session after 15 minutes. The first session will be a practice session.

Thank you for your cooperation.

Good Luck!

Appendix 7: Map Task - ASR tips for Hearing Participants

Use 'careful' speech.
Speak in short phrases and sentences.
Try not to say 'um'.
At the end of a sentence, say 'period' or 'question mark'.
Between new topics or ideas, say 'new paragraph' or press the 'return' key.

Appendix 8: Session Feedback Questionnaire: Deaf Participants

ASR Study — Deaf Participants' Feedback Questionnaire

Participant ID: ________________ Condition: ________________ Date: ______

1. How easy or difficult was it to understand the other person in this conversation?	Very difficult									Very easy		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
2. I asked my partner to repeat:	Very often									Not at all		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
	Not at all								Completely N/A			
3. I relied on speech reading	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
4. I relied on hearing	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
5. I relied on the text	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
6. The text was:	Very difficult to follow								Very easy to follow			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
7. The text had:	Many errors								Few errors			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
8. My partner was:	Very difficult to speech-read									Very easy to speech-read		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N/A**	
9. During the conversation, I felt:	Very uncomfortable							Very comfortable[19]				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
10. I would try this mode of communication again.	Strongly disagree							Strongly agree				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		

Comments:

[19] For *Session 1: Conversation* data collection, the valence for #9 was printed incorrectly (very comfortable =1 vs. very uncomfortable = 10). The mistake was discovered and corrected before *Session 2: Map Task* data collection and so did not affect data presented in this dissertation.

Appendix 9: Session Feedback Questionnaire: Hearing Participants

ASR Study — Hearing Participants' Feedback Questionnaire

Participant ID: _________ Condition: _________________ Date: ________

1. How easy or difficult was it for you to communicate in this conversation?

Very difficult									Very easy
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

2. I changed my normal way of speaking

Very much									Not at all
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Comments:

3. My partner was:

Difficult to understand									Easy to understand
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Answer only if you used Speech Recognition this time:

4. I corrected the text:

Very often									Not at all
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

5. The system made errors:

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Comments:

6. During the conversation, I felt:

Very uncomfortable									Very comfortable[20]
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Comments:

7. I would try this mode of communication again (to converse with a deaf person)

Strongly disagree									Strongly agree
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Comments:

---

[20] For *Session 1: Conversation* data collection, the valence for #9 was printed incorrectly (very comfortable =1 vs. very **uncomfortable** = 10). The mistake was discovered and corrected before *Session 2: Map Task* data collection.

Appendix 10: Final Feedback Questionnaire: Deaf Participants

Participant ID: _______ Date: _____

Please rank the communication conditions in terms of communication efficiency (based on both the conversation and the Map Task sessions)

1 = most efficient 5 = least efficient

___ Face-to-face
___ ASR only
___ ASR plus face-to-face
___ CART only
___ CART plus face-to-face

What improvements would you suggest for the ASR?

Appendix 11: Final Feedback Questionnaire: Hearing Participants

Participant ID: _______ Date: _____

Please rank the communication conditions in terms of communication efficiency (based on both the conversation and the Map Task sessions)

1 = most efficient 5 = least efficient

___ Face-to-face
___ ASR only
___ ASR plus face-to-face
___ CART only
___ CART plus face-to-face

What improvements would you suggest for the ASR?

Was your practice time with the ASR sufficient? Explain.

If you had more time to practice on the ASR, do you think you could have communicated with your partner more efficiently?

Appendix 12: Landmark Scoring Notes

Use the maps with the overlaying grids while scoring.

General Process of Scoring:

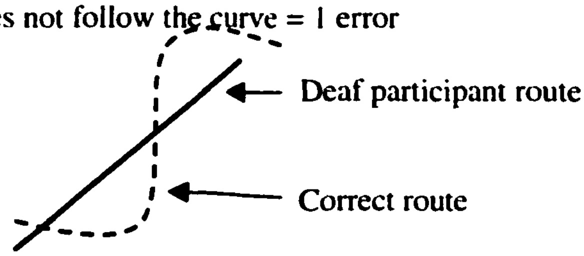
Take off one point for each error.

Count the number of relevant landmarks. In other words, count the number of landmarks that are used to navigate the route.

Calculate a ratio for the *# of errors* : *# of landmarks*. If the Map Task is incomplete, landmarks not reached are scored as errors.

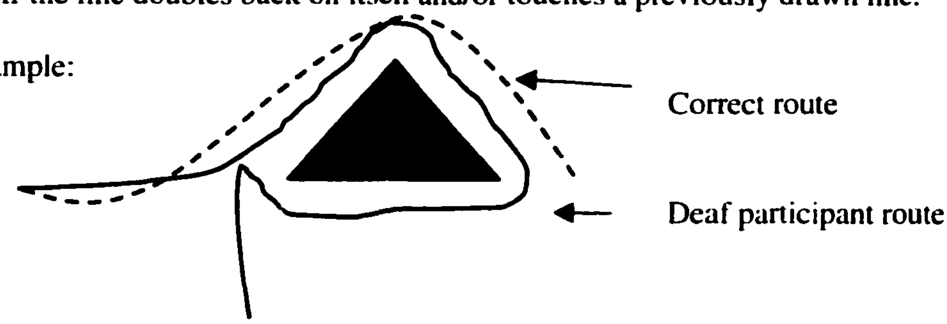
Wrong errors include:

- Blatant errors
- If the line goes through the middle of a landmark = 1 error
- If the line does not follow the curve = 1 error
  Example:

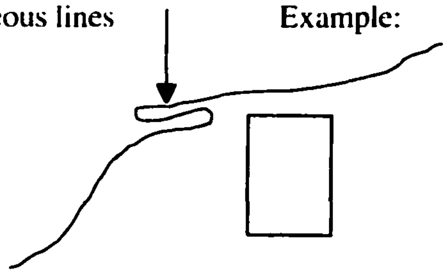


- If the line does not go along the proper side of the landmark
- If the line doubles back on itself and/or touches a previously drawn line.

Example:



Extraneous lines   Example:



Specific Scoring Rules for each map:

Map 1

Waterfall curve properly

Map 3

Canal/South Bay issue – If the line goes too far from where Canal should be, and the line just drops from that point to go to South Bay, South Bay can be considered an error also. So, take off for both.

Map 4

Lion Country and Limestone Cliffs should be scored as one "package." Count as one error.

East Lake and Lion Country – Take off for both. See Canal South Bay Issue.

Map 5

Missionary Camp – If the line does not go down to left side of missionary camp, take off a point.

Furniture Store – Take off a point if the line seems to double back a bit.

Appendix 13: Transcription Editing Notes

1) Order of Transcription Editing

a) Open file: eg: Pair 5 (4) CF1.doc
b) Save in subdirectory Pair 5 Edit and save as Pair 5 (4) CF1 edit.doc (add '**edit**' to filename).
c) Change to normal view.
d) Turn on **track changes**! *(TOOLS → TRACK CHANGES)*
e) Add Pair number (e.g. **4D**, **4H**) to talker column [*see (2)(b)]. Use ***find and replace***.
f) Copy summary data, work time, info from previous transcript and update
g) Open a file called ***'ASR/CART/FTF Accuracy Notes.doc'*** and keep a record of comments & observations for each dialogue.
h) **1st pass** - Listen to dialogue and add time counter #'s and calc total time, convert to minutes (sec/60).
   a. **Start time =** 1st utterance about the task. e.g. "ok the start is ..." (the 1st sentence may not have audio on some recordings, but count them anyway!)
   b. **End time** = end of last utterance about the task – no comments about performance, etc. e.g. "Yay, We're done."
i) **2nd pass** – Edit speech output (add *hesitations*, {*extraneous sounds*}, <*interruptions/*).
j) **3rd pass** – Edit ASR/CART/typing output.
k) **4th, 5th pass** – Listen 2 more times for final edit of speech and ASR/CART/typing output.

2) The Columns

a) **Count** – Clock marker (time in seconds) shows approximate beginning of utterance for each row. (Note: If we want real accurate pause and time measures we will need to convert to WAV files.)
b) **Talker** – Make sure it says Pair number and D (for deaf participant) or H (for hearing participant). (e.g. 6H) [D for A, H for B]
c) **Utterance**
   i) This column should document **EVERY** participant production.
   ii) Edit to be as ACCURATE as possible.
   iii) Include all hesitations, repetitions, dysfluencies, coughs, laughter, throat clearing. (Important: ASR may have transcribed coughs or throat clearing.)
   iv) Include participant typing – they are PRODUCTIONS. Use blue ink. (See Pair 1 A3.)
d) **ASR/CART/Typing Output**
   i) Use Copperplate Gothic Bold font. Use blue ink for typing output.
   ii) For ASR, compare with ASR AVI files for typed utterances.
   iii) **DO NOT** add punctuation (or anything else) to ASR output, unless speaker typed such punctuation, or uttered such commands as 'question mark' or 'period.'

iv) Use BLUE ink to indicate **participant typing**, including punctuation marks that are typed. This should also be copied and pasted into the Utterance column.
v) Use PINK ink to indicate commands misunderstood by ASR that were written as part of dialogue.
vi) Use GREEN ink to indicate dialogue misunderstood by ASR and interpreted as a command. Also include how the problem was fixed. Eg: Pair 2 (4) AF5, line 21
vii) Use OUTLINE to indicate anything that is erased, either by the ASR command "scratch that" or physically done by the participant.
viii) Using ORANGE ink, write "(inserted!)" to indicate that text has been inserted into already existing ASR output – this would typically be a result of H speaking after having left the ASR cursor within printed text, usually for correction purposes. Eg: Pair 5 (5) AF4, line 866.
ix) In continuation of (viii), use *ARIAL NARROW-italicized* for the existing text into which the new ASR output is inserted.

3) Utterance Types

a) **Unintelligible utterances –** First, try REALLY hard to understand, but don't make something up! As a last resort, transcribe as "(unintelligible)." This is tallied as one word in the word count. These most often occur during speech overlaps. There are instances in which the unintelligible utterance is a dysfluencies or extraneous sound.

b) **Contractions** – should be transcribed the way they were produced – not separated – unless that's how they were said. However, these should be changed to reflect ASR output. (E.g. **there's, it's, who's, can't, don't**). This rule is for the sake of consistency across transcribers. BUT phrases that are often 'mashed' together in speech should not be written that way:

Write:	going to	for:	gonna
Write:	have to	for:	hafta
Write:	want to	for:	wanna

However, there is an instance where the ASR transcribes wanna.

Pair 3 (6) AF2		Speech	ASR
584	6H	Now you <u>wanna</u> draw a	Now you <u>wanna</u> draw a

c) **Hesitations/Incomplete words**

i) Incomplete word: Use Phonetic transcription and put in Square brackets e.g. [əb] above You can find the symbols under Insert → Symbol. Symbols are in the Lucinda Sans Unicode font. [ə] Subset: IPA extensions: [æ] Subset: Latin-1.
ii) Complete word (part of compound) – don't need to transcribe:e.g. Go clock, clockwise.

d) **Meaningful non-words –** For the most part, use CART stenographer's transcription – no need for total phonetic accuracy.

i) Huh? = What?
ii) Uh-huh, Mhm, Yeah = Yes
iii) Uh-uh, M-m = No

e) **Extraneous sounds** – laugh, cough, clear throat (important if transcribed by ASR). IMPORTANT – to transcribe hesitations for 'B' (hearing person) because ASR transcribes them, which can affect accuracy scores. Therefore you must also transcribe for deaf - otherwise it will appear that the hearing person uttered more words. Code this way: {cough}; {laugh}; {clearthroat}.

f) **Anita's interjections** – If Anita interjects during a map session, highlight those exchanges in turquoise. Include in total TIME, but exclude exchange between Anita and participant(s) from coding and word counts. (See Pair 1 AF2.)

g) **Overlapping Speech/Interruptions** – Absolute accuracy is only possible with WAV files.

   i) Mark overlaps any time you hear them.

   ii) Include ASR commands and typing (when typing overlaps with the other person speaking). e.g. Pair 2 A3 Note yellow highlight and brackets:

2B	Yes. I do.
2H	<Yes./
2B	Period.> New paragraph.

   iii) **Do not put a space between the symbols '< / >' and the utterances.** Otherwise MS Word will count the symbols as words when doing a word count. In other words, put them in as follows: <abc/ xyz>

   iv) **Do NOT code as overlaps** – Extraneous sounds {cough} that overlap with speech.

   v) Later, I will come up with a coding scheme for overlaps (back channels) vs. interruptions. I've also noticed that many interruptions (especially in FTF) do not have overlapping speech.

h) **Erased Words** – For words that have been erased with 'scratch that' or 'delete/backspace, use FONT 'OUTLINE' (See Pair 4 session 3).

i) Nonverbal gestures (head nod) – If the partner responds to them, code for turns, but not for transcription accuracy errors. [The TURN is accounted for, but there was no ERROR.]

Pair 1 CF5				
After 317	1H	(1H nods)		

The following is slightly different… The mouthed words (line 36) should be marked as a separate turn and counted in the word count (because 3D responded to those words).

Pair 3(4) F4 **FACE-to-FACE**		
35	3D	Go down three more inches?
	3H	Three more inches. (Can lipread, but can't hear 3H saying this.)
	3D	Okay.

4) *Word Count*

a) Do a word & turn count for each of the following:

- Total number of words
- Total number of words spoken by the deaf participant
- Total number of words spoken and typed by the hearing participant
- Total number of words spoken by the hearing participant, excluding typed words

- **STUCKLESS**: Total number of words spoken by the hearing participant, excluding typed words, commands and extraneous sounds.
- **COMMUNICATIVE**: Total number of words spoken by the hearing participant, excluding typed words, commands, extraneous sounds and dysfluencies.
- (ASR) Total number of words from ASR output.
- (ASR) Total number of words from ASR output, excluding typed words.
- (CART) Total number of words from CART Output.
- (CART) Total number of words from CART Output, excluding commands and extraneous sounds. (**STUCKLESS**)
- (CART) Total number of words from CART Output, excluding commands, extraneous sounds, and dysfluencies. (**COMMUNICATIVE**)

b) Create a table at the end of the document to input data collected from word counts.

	WORD	TURN	
TOTAL			
DEAF			
HEAR			
HEAR (no type)			
HEAR (Stuckless)			No commands, extraneous sounds
HEAR (Communicative)			No commands, extraneous sounds, disfluencies

For ASR:

ASR			
ASR (no type)			
ASR (Stuckless)			No commands, extraneous sounds
ASR (Communicative)			No commands, extraneous sounds, disfluencies

For CART

CART HEAR			
CART (Stuckless)			No commands, extraneous sounds
CART (Communicative)			No commands, extraneous sounds, disfluencies

c) For doing word/turn counts for ASR:

i) Save file as SCRAP or something to use instead of the original document (just in case).

ii) Delete header.

iii) Delete first 2 rows. *Keep in mind which heading belongs with which column.

iv) Select what is now the first column (formerly the TALKER column), and sort in descending order so that all the H utterances at the beginning of the document. *(TABLE → SORT)*

v) Delete anything highlighted in turquoise (Anita's exchanges with the participant), all paragraph markings, and any notes you may have made in this document.

vi) Cut columns formerly known as TALKER and UTTERANCE.

vii) Do a word/turn count for total number of words from ASR Output, including typed words. *(TOOLS → WORD COUNT)*. Get the turn count from the number of paragraphs.

viii) Delete typing, and do a word/turn count for total number of words from ASR Output, excluding typed words.

ix) Undo cut made to TALKER and UTTERANCE column.

x) Delete ASR Output column.

xi) Cut TALKER column.

xii) Do a word/turn count for total number of words.

xiii) Undo cut made to TALKER column.

xiv) Select all rows containing H utterances (which should be at the beginning of the document), cut, and paste in a new document.

xv) In this new document, delete the TALKER column.

xvi) Do a word/turn count for total number of words spoken and typed by the hearing participant.

xvii) Delete typing, and do a word/turn count for total number of words spoken by the hearing participant, excluding typed words.

xviii) Delete commands and extraneous sounds, and do a word/turn count for total number of words spoken by the hearing participant, excluding typed words, commands, and extraneous sounds (**STUCKLESS**).

xix) Delete dysfluencies (hesitations, um's, er's, uh's, false starts, etc.), and do a word/turn count for total number of words spoken by the hearing participant, excluding typed words, commands, extraneous sounds and dysfluencies. (**COMMUNICATIVE**).

xx) Go back to SCRAP (the document which now only has D utterances), and delete the TALKER column.

xxi) Do a word/turn count for total number of words spoken by the deaf participant.

d) For doing word/turn counts for CART:

i) Save file as SCRAP or something to use instead of the original document (just in case).

ii) Delete everything except the TALER, UTTERANCE, and CART OUTPUT columns. Delete anything highlighted in turquoise or typed in gray.

iii) Sort by TALKER column.

iv) Copy UTTERANCE column (which includes both H & D) and paste into a new document.

v) Do a word/turn count for total number of words.

vi) Discard the new document (you won't need it anymore).

vii) Cut all the D-rows and paste them into a new document.

viii) Delete TALKER & CART OUTPUT columns.

ix) Do a word/turn count for total number of words spoken by the deaf participant.

x) Discard this document.

xi) Copy H's utterances and paste into a new document.

xii) Do a word/turn count for total number of words spoken by the hearing participant, excluding typed words, commands, and extraneous sounds (**STUCKLESS**).

xiii) Do a word/turn count for total number of words spoken by the hearing participant, excluding typed words, commands, extraneous sounds and dysfluencies. (**COMMUNICATIVE**).

xiv) Discard this document.

xv) Copy CART OUTPUT for H and paste into a new document.

xvi) Do a word/turn count for total number of words from CART Output.

xvii) Do a word/turn count for total number of words from CART OUTPUT, excluding commands and extraneous sounds. (**STUCKLESS**).

xviii) Do a word/turn count for total number of words from CART OUTPUT, excluding commands, extraneous sounds, and dysfluencies. (**COMMUNICATIVE**).

Appendix 14: Number of Words per dialogue

	CONDITION						
PAIR	**FTF Only**	**CART-FTF**	**CART Only**	**ASR-FTF**	**ASR Only**	***mean***	***(s.d.)***
1	1007	904	836	777	913	887.4	86.70
2	1789	1094	1403	1539	1403	1445.6	251.94
3	717	1201	1231	811	1054	1002.8	230.46
4	609	1070	601	1281	1182	948.6	322.43
5	777	954	536	1127	1072	893.2	240.53
6	740	771	768	727	874	776.0	57.86
mean	939.8	999.0	895.8	1043.7	1083.0	1013.2	
(s.d.)	436.15	153.56	348.10	326.85	192.74	255.89	

Appendix 15: Transcription Accuracy Coding Manual

(Adapted from **NTID Test of ASR Readability** in Stuckless, R. (1998). Recognition means more than getting the words right: beyond accuracy to readability. *Speech Technology Magazine*. October/November.)

SHARED CODING ITEMS

1) SUBSTITUTION (BRIGHT GREEN)

a) Regular (S): single spoken word transcribed by ASR or CART as another word.

Cnt.	Tkr.	Utterance	ASR	NY Code
188	1H	And curves up to the right	In curves up to the right	S=1
Pair 1 (4) A3		(Syllable added – still simple substitution)		
397	1H	Are you at the line of south bay?	Are you at the bowline of south bay	S=1
Pair 2 (4) AF5		(One word transcribed as 2 words)		
382	2H	Now draw the line downward about 7:00. Period.	Now draw the line down word about 7:00.	S=1
Adapted from Pair 2 (4) AF5		(String of words transcribed as a single word):		
674	2H	So just to make sure we're correct, the line goes around white mountain, continues around white mountain clockwise so that the line is between furniture store and white mountain.	Statistics sure we're correct the line goes around White Mountain continues around White Mountain clockwise subline is between furniture store and White Mountain	S=4 S=4
Pair 4 (4) A5		(Remember to score **DELETED** words for accuracy)		
1	4H	Okay. Your starting point should be an inch above the missionary camp. Select <learn starring client./	Okay you ~~learn starring client~~ should be an inch above the missionary can	S_{end}=1 A=1 S=2 S=1

i) ***Subcategories of Regular*** (do not necessarily need to code so specifically on first pass)

Homonym (S_{hom}) – single spoken word transcribed as homonym.

Pair 3 (6) AF2				
27	6H	Go down two inches	Go down to inches	S_{hom}=1

Ending (S_{end}) – single spoken word transcribed with change in ending (i.e., any additions or omissions to the ending of the spoken word), including tense, possessive, number, contraction of two words, part of speech.

Pair 6 (6) A3				
0	6H	The finish mark is	The finish marked is	S_{end}=1
Pair 1 (4) A3				
383	1H	And finishes to the left	And finish is to the left	S_{end}=1

But note:

Pair 5 (3) A2		2 words transcribed as 1 word.		
813	5H	Finish. Finish is -- the finish is about three inches	Finish finishes the finishes about three inches	S=2 S=2

ii) Disfluency (S_{dis})

Disfluent word(s) or filler(s) incorrectly recognized by ASR/CART as a spoken utterance. (Dysfluencies include false starts, fillers, repairs, and repetitions.) We understand that Dragon Natural Speak is not programmed to transcribe disfluencies at this time, so it cannot transcribe dysfluencies correctly.

Pair 6 (6) A3				
37	6H	Now, go [f] – Sorry	Now go for a Sorry	S_{dis}=1
Pair 2 (4) AF5		Unintelligible Disfluency		
735	1H	(unintelligible$_{dis}$)	The DC to	S_{dis}=1
Pair 4 (3) A2		Substitution Disfluency – ASR transcription is part of a larger word		
1	4H	So your [s] start	So vigorous start	S=1 S_{dis}=1

iii) Extraneous (S_{ext})

Any utterance that does not have a communicative purpose and is not a disfluency, e.g. cough, laugh, a clearing of throat, conversation-partner noise, room noises.

Pair 4 (4) A5				
19	4H	{Clearthroat}	Between	$S_{ext}=1$
Pair 4 (3) AF2 Deaf person speaking loudly – ASR transcription occurred				
303	4D	Oh, wait, wait. No, I do have. I am I am (unintelligible).	All	$S_{ext}=1$

OMISSION (PINK)

Regular (O) – single spoken word that is uttered but not recognized by ASR/CART at all.

Pair 4 (4) A5				
19	4H	Do you have	Do have	O=1

Disfluency (O_{dis}) – disfluent word(s) or filler(s) not recognized by ASR/CART at all. (Disfluencies include false starts, fillers, repairs, and repetitions.)

Pair 4 (4) A5				
390	4H	(Unintelligible$_{dis}$)>		$O_{dis}=1$

Extraneous (O_{ext}) – any omitted utterance that does not have a communicative purpose and is not a disfluency such as a cough, laugh, a clearing of throat, etc.

Pair 4 (4) A5				
1	4H	{Tsk} Okay	Okay	$O_{ext}=1$

ADDITION (LIGHT GRAY)

Regular (A) – transcription that cannot be linked to any utterance.

Adapted from Pair 3 (6) AF2				
27	6H	Down two inches.	Down two the inches	A=1

ASR-SPECIFIC CODING ITEMS

These concern ASR commands like *Period*, *New Paragraph*, and *Scratch That*.

SUBSTITUTION (BRIGHT GREEN)

Dialogue-C (S_{dia-C}) – dialogue misinterpreted by ASR as a command.

E.g. Correct. [*Word in dialogue*] → Correct [*ASR Command*]

Pair 2 (4) AF5 ASR misinterpreted dialogue as a command. Following words are omissions.				
462	1H	Correct. Now do you see hidden steps?	**ASR misinterpreted the "correct" as an ASR command, rather than a word in the dialogue. ASR was looking for words to correct and did not transcribe the utterance that followed.**	$S_{dia-C}=2$ O=6
Pair 2 (4) AF5 Same as above.				
470	1H	Correct. Now stop. Now go counterclockwise above and around hidden steps.	(ASR command: correct 'steps')	$S_{dia-C}=1$ O=10
Adapted from Pair 3 (6) AF2 "your line" ASR understood as [command: new line]				
282	6H	Okay. Now, from your line you're going to draw a diagonal line between the rocks and the stone creek.	Okay now from You're going to draw a diagonal line between the rocks and the stone creek.	$S_{dia-C}=2$

Command-D (S_{com-D}) – a command misinterpreted by ASR as dialogue.

Pair 1 (4) A3 2-word commands (e.g., *new paragraph*) are counted as 2 errors.				
188	1H	New paragraph. New paragraph.	Depend	$S_{com-D}=2$
Pair 4 (4) A5				
64	4H	New paragraph. Scratch that.	Repair	$S_{com-D}=2$
Pair 4 (4) A5 Move to end of line = Routes and line; scratch that = 'z'				
19	4H	Move to end of line. Scratch that. Move to end of line. Scratch that.	Routes and line Move to end of linez	$S_{com-D}=5$ $S_{com-D}=5$ $S_{dia-C}=2$

Extraneous-C (S_{ext-C})

There was one occurrence of a Substitution of an Extraneous sound for a Command. This was coded as $S_{ext-C}=1$.

Pair 5 (3) A2

451	5H	{tsk}. Go south of the rapids.	**ASR went into correction mode and did not transcribe the following utterance.** ***Similar to S_{dia-C}, Pair 2 (4) AF5, line 462 above.**	$S_{ext\text{-}C}$=1

OMISSION (PINK)

Command (O_{com}) – any commands that are omitted (i.e., ignored) by ASR.

Pair 2 (4) AF5				
570	1H	Period. Now draw	Now draw	O_{com}=1

CART-SPECIFIC CODING ITEMS[21]

SUBSTITUTION (BRIGHT GREEN) Turn-D ($S_{turn\text{-}D}$) –

Mistyping of key strokes for new speaker's turn appears as part of dialogue in the previous turn. In these instances, the utterance is usually transcribed – it's just that the CART Reporter failed to indicate a new turn/speaker.

"Are char" is coded below as $S_{turn\text{-}D}$ because it was meant to indicate a new speaker but was typed incorrectly.				
Pair 2 (5) C4				
The following was taken straight from CART Unedited Pair 2 (5) C4: **SPEAKER 2D** I'm at the top of waterfall. Are char draw a line to the right about one inch				
561	2D	I'm at the top of waterfall.	I'm at the top of waterfall.	
563	2H	Okay. Draw the line to the right about one inch, so now	Are char draw a line to the right about one inch, so now	$S_{turn\text{-}D}$=1 O=1 S=1
"When EL mel" is coded below as $S_{turn\text{-}D}$ because it was meant to indicate a new speaker but was typed incorrectly.				
Pair 1 (6) C4				
376	1H	Okay, follow the …	When EL mel okay, follow the …	$S_{turn\text{-}D}$=1
"Gentlemen JOE" is coded below as $S_{turn\text{-}D}$ because it was meant to indicate a new speaker but was typed incorrectly.				
Pair 5 (4) CF1				
424	5D	…the right side of the pyramid.	… the right side of the pyramid. Gentlemen **JOE** I want you on the left side. Go to the left side of the pyramid.	
427	5H	I want you on the left side.	Gentlemen **JOE** I want you on the left side.	$S_{turn\text{-}D}$=1

Turn (S_{turn}) – CART Reporter's failure to correctly specify new speaker, although a new turn was indicated.

Note "A" is coded as S_{turn} because although a new speaker was indicated, the new speaker's name was typed incorrectly.: ("A" was mistyped to indicate new speaker)				
Pair 6 (5) CF1				
The following was taken straight from CART Unedited Pair 6 (5) CF1: **SPEAKER 6H** Okay. Do you have a pyramid at the top of your page a little more to the right. A. Top of the page and slightly to the right is a pyramid, right.				
453	6H	Okay. Do you have a pyramid at the top, top of your page a little more to the right?	Okay. Do you have a pyramid at the top of your page a little more to the right.	O_{dis}=1
464	6D	Top of the page and slightly to the right is a pyramid, right.	**A.** Top of the page and slightly to the right is a pyramid, right.	S_{turn}=1
In the following example, "Q" is coded below as S_{turn} because although a new speaker was indicated, the CART Reporter failed to correctly input the speaker's name, and was thus included in the text. (*See Pr2 (6) CF1, lines 387-398)				
Pair 2 (6) CF1				
387	2D	White mountain. (2D nods)	A. White mountain.	
389	2H	Okay. Draw the line …	**Q.** Okay. Draw the line …	S_{turn}=1

OMISSION (PINK)

---

[21] All instances of CART Turn-Substitution in Map Task dialogues are noted here.

Turn (O_{turn}) – Omission of a new turn (or, next speaker not indicated). In these cases, the CART Reporter does not necessarily transcribe the utterances.

Pair 4 (2) CF3				
160	4H	Okay. Draw the line to the right so that it's parallel with the end of the cliffs.	Okay. Draw the line to the right so that it's parallel with the end of the cliffs. <u>So about an inch to the right.</u>	
	4D	Okay.>		
172	4H	So about an <inch/ to the right.	So about an inch to the right.	O_{turn}=1

In the following example (and in many others) the text <u>underlined and in gray</u> was once a part of the utterance in that cell. However, after listening to the avi files again, utterances by H were detected and inserted as heard into the transcript.

Pair 2 (6) CF1				
36	2D	… you're the one who has control of <this with the map/	… you're the one who has control of this with the map, <u>if you're saying no, then it's gone. It's like it never existed here, okay? That makes it a little easier for me.</u>	
42	2H	Right, right.>		O_{turn}=1 O=2
	2D	If you're saying no, then it's gone.	if you're saying no, then it's gone.	
46	2H	Okay.		O_{turn}=1 O=1
	2D	It's like it never existed here, <okay?/	It's like it never existed here, okay?	
48	2H	Okay.>		O_{turn}=1 O=1

ADDITION (LIGHT GRAY)

Punctuation (A_{punct}) – If punctuation stands alone, it is counted as a word during a word count. It will be excluded from the Stuckless/Communicative scores. This is the only incidence in the entire Map Task corpus.

Pair 2(5) C4				
608	2H	… fallen rocks.	… fallen rocks. ,	A_{punct}=1

Don't worry about coding/scoring or doing a word count for CART output of <u>deaf</u> participant. It doesn't really matter anyway because these are only included in the Total count.

Transcription of Overlaps – it's more important to keep a turn together than to break it up by the occurrence of an overlap. Note also that overlaps are not always interruptions!

Pair 1 (6) C4				
The following has been changed from:				
299	1D	Fallen rocks? It's way over to the right. It's almost on the right side, right, <right edge/	Fallen rocks? It's way over to the right. It's almost on the right edge of the paper.	O_{dis}=3
304	1H	Yes.>	Yes. You're going to go all the way diagonally down and around the bottom of the fallen rocks.	
	1D	of the paper.		
305	1H	You're going to go all the way diagonally down and around the bottom of the fallen rocks.		
Simplified to:				
299 listen	1D	Fallen rocks? It's way over to the right. It's almost on the right side, right, <right edge of the paper./	Fallen rocks? It's way over to the right. It's almost on the right edge of the paper.	
304	1H	Yes.> You're going to go all the way diagonally down and around the bottom of the fallen rocks.	Yes. You're going to go all the way diagonally down and around the bottom of the fallen rocks.	

Transcription of Fillers – Do no use brackets "[ ]": um = um uh = uh er = er.

Like is a filler. Count like as an um, or uh = disfluency.

Pair 6 (5) CF1			
After 295	6H	Mhm. Like, so it goes …	Uh-huh. So it goes …

The "or" for the "er"" is a Substitution$_{disfluency}$ error.

Pair 4 (2) CF3				
364	4H	an inch to the left of… the canal, **<u>er</u>**, of the bakery I have a canal…	an inch to the left of… the canal - - **<u>or</u>** of the bakery I have a canal.	S_{dis}=1

			...	
If we were unsure of the utterance transcription, we defaulted to CART output. We weren't sure if 5H really said "or" or a filler, like *er* and *uh*.				
Pair 5 (2) C3				
299	5H	To the edge of... the paper, or, to edge of the map.	To the edge of... the paper or to the edge of the map.	

Affirmatives/Negatives – Yeah, Yes, Yup, Mhm, Uh-huh are all have essentially the same meaning. Discrepancies in ASR/CART transcription do not get penalized as substitution errors. (This is mainly a CART issue.)

Pair 1 (6) C4				
98	1H	Mhm.	Uh-huh.	
Pair 2 (6) CF1				
235	2H	Yeah.	Yes, and it's closer to footbridge ...	
Pair 4 (2) CF3				
34	4H	Mhm.	Uh-huh.	
Pair 6 (5) CF1				
After 295	6H	...Mhm. Like, so it goes like ...	Uh-huh. So it goes like ...	O=1
Note: In this case, however, the message of "oh" is slightly different than "um."				
Pair 3 (5) C5				
779	3H	Um. Okay. The gazelles should ...	Oh. Okay. The gazelles should ...	$S_{dis}=1$

Uses of "--" by CART reporter

There were times when the CART reporter has "--" in place of an extraneous sound. Leave it alone, but remember to delete them before doing a word count.

Pair 4 (2) CF3				
276	4H	so it's about {**groan**} down and to the right	so it's about -- down and to the right	$O_{ext}=2$ $O_{ext}=1$

This is one is a little different because the "--" doesn't directly translate into the "er":

Pair 4 (2) CF3				
364	4H	an inch to the left of the canal, er, of the bakery I have a canal.	an inch to the left of the canal -- or of the bakery I have a canal.	$S_{dis}=1$

Used to indicate the speaker's hesitation/pause.

Pair 1 (6) C4			
250	1H	Do you have– I have a waterfall two inches above my railroad crossing.	Do you have -- I have a waterfall two inches above my railroad crossing.
Pair 3 (3) CF3			
559	3H	Because my wheat fields are, let's see, an inch to the left of the south bay.	Because my wheat fields are -- let's see -- an inch to the left of the south bay.

Used to indicate an interrupted utterance.

Pair 2 (5) C4				
388	2H	{k n}.> continue the line to the left of picnic site and {g bl}, and <draw a line/	Continue the line to the left of picnic site and draw a line --	$O_{dis}=1$ $O_{dis}=2$
396	2D	To the left> of picnic site.	To the left of picnic site.	

Used to indicate a disfluency. In the following example, instead of putting in the disfluency, the CART reporter substituted it with "--".

Pair 2 (5) C4				
652	2H	Make sure the line is between <lion country and lime[st]/	Make sure the line is between iron country --	S=1 $O_{dis}=2$

Used to indicate a pause in H's utterance.

Pair 3 (5) C5			
50	3H	Okay. You want to draw gorillas there	Okay. You want to draw gorillas there --
	3D	<Gorillas./	Gorillas.
53	3H	About> an inch from the bottom of the paper	-- about an inch from the bottom of the paper

Interruption – [12/21/00] Agreed that since a "mental interruption" can be seen as causing a disfluency, an argument could be made for considering an interrupted utterance as a disfluency as well, so count it as a disfluency. A disfluency that appears to be caused by an interruption is still coded as a disfluency.

For the following example, note that the omitted words are coded as a disfluencies.

The first part of an interrupted (disfluent) utterance is considered the error. This will decrease the number of errors in the communicative count.

The following has been changed FROM:

Pair 3 CF3				
268	3H	No. I don't <have./		O=4
	3D	(Unintelligible)>		
	3H	I don't have a green bay.	No. I don't have a green bay.	O=1

Changed TO:

Pair 3 CF3				
Right after 268	3H	No. I don't <have./	No.	O_{ih}=3
	3D	(Unintelligible)>		
	3H	I don't have a green bay.	I don't have a green bay.	

Contractions – Change contractions in the Utterance column to reflect or agree with ASR/CART Output column.

The following has been changed FROM:

Pair 3 (5) C5			
312	3H	**It'll** be the left side …	It will be the left side …

Changed TO:

Pair 3 (5) C5			
312	3H	**It will** be the left side …	It will be the left side …

The following has been changed FROM:

Pair 3 (3) CF3			
363	3H	… The **line's** going to stop …	… the line is going to stop …

Changed TO:

Pair 3 (3) CF3			
363	3H	… The **line is** going to stop …	… the line is going to stop …

Compound words – For differences in compound words, change what we've transcribed while listening/watching the videos to match the ASR/CART transcription.

Just changed the Utterance column to reflect the ASR/CART.
Note: ASR anomaly: Left-hand and right hand.

Pair 2(5) C4			
563	2H	… waterfall.	.. water fall.
563	2H	… water fall.	.. water fall.

Pair 1 (2) AF2	
wagon wheel	wagonwheel
stone creek	stonecreek
gold mine	goldmine
whitewater	white water
Pair 5 (3) A2	
southwest	south west
southeast part	south eastport

Inversions – Code them as substitutions.

Pair 1 (6) C4				
If this were H's utterance, it would be coded as S=2				
28	1D	… Do I go to the left …	… I do go to the left …	
Pair 1 (5) CF5				
32	1H	Straight across> and around …	Straight around and across …	S=1 S=1

Interruptions that are dysfluencies

Pair 1 (6) C4				
373	1H	[f]>		O_{turn}=1 O_{ih}=1
Pair 2 (5) C4				
After	2H	[s]>		O_{turn}=1

332				O_{dis}=1=

Repetitions – Repetitions are produced for emphasis, so do not score as disfluencies.

Pair 4 (2) CF3				
692	4H	Like it's halfway up, halfway up, but	Like it's halfway up, but	O=2 (Repetition)
Pair 5 (2) C3				
260	5H	Right. Uh. No, no, no. If you	Right. No, no, if you	O_{dis}=1 O=1 (Repetition)
Pair 2 (6) CF1				
133	2H	Okay, yes. Yes.> <Yes./		O=4

PROBLEMS

Questionable Addition

6H prolonged the [n] sound at the end of the word begin – this was interpreted by the ASR as "in". Counted as A=1.				
Pair 6 (6) A3				
0	6H	Now we are ready to begin [n].	Now we are ready to begin in	A=1
What we did: Deleted prolongation from utterance column.				

Human error when attempting ASR Commands

What do you do for this: When she said "lathes" the second time in line 562, she meant "select lathes. And when she said, "delete" she meant to erase it, thereby giving a command. Do you count those as commands, and therefore, not include them from the word count?

Pair 4 (4) A5				
562	4H	Draw the [n], the line one inch and a half to the left. Lathes.	Draw the in the line one inch and a half to the lathes ~~lathes~~	S=1
	4D	To the left.		
	4H	Select lathes.		
574	4D	<Lathes/		
	4H	Delete>	Delete	
	4D	Is left.		
577	4H	Scratch that, scratch that. Scratch that. New paragraph.		
What we did: Attempts at commands (misuse of software) are NOT commands [because we can't know for certain what the speaker's intent really is]. Included attempts in Stuckless/Communicative word counts.				
Should the first "Question" be included in the word count?				
Pair 4 (4) A5				
193	4H	So you should be about an inch to the left of the valley, right? Question. <Question mark./	So you should be about its an inch to the last of the Valley right?	A=1 S=1 O_{dis}=1
What we did: 'Question' is a disfluency. Scored as disfluency. Excluded when doing Communicative word count.				

Disfluency

We think the CART Reporter transcribed /k / take as *continue*. How to score???				
Pair 2 (5) C4				
398	2H	And [k] take the line to the right going below picnic site.	Continue the line to the right going below picnic site.	S_{dis}=1 O=1
Is *Like* a disfluency?				
Pair 6 (5) CF1				
After 295	6H	[serk]> Mhm. Like, so it goes like up and around. [hand motion]	Uh-huh. So it goes like up and around.	O_{dis}=1 O_{dis}=1 Is "Like" a disfluency? Count like as an um, or uh. = disfluency - abh

Transcription error occurs in the other participant's utterance. The following example was not scored because it corresponded with D's utterance. However, please note that the error actually occurred in H's utterance. In the following example, since there was no paragraph mark, it was scored as S_{turn}. If a new line or new paragraph was indicated by a paragraph mark, I marked as a new turn, and did not score as an S_{turn} error. (See Pair 2 (6) CF1, line 385-398)

Pair 1 (6) C4				
34	1H	… to the very edge of the page.	… to the very edge of the page. A Lou okay.	$S_{turn-17}=1$
39	1D	Okay.	A Lou okay.	

NOTES OF INTEREST (CART)

Pair 2 (6) CF1		DEAF SPEAKER – does not get coded	
292	2D	Go underneath <polluted stream? Under polluted/	Go underneath polluted stream. Under **owe looted** stream.
Pair 2 (6) CF1		DEAF SPEAKER – does not get coded	
393	2D	I'm going to the left of white mountain.	**Ing** if to the left of white mountain.
Pair 2 (6) CF1		DEAF SPEAKER – does not get coded	
457	2D	And the left of remote village.	**SZ SUZ** and the left of remote village.

Speaker meant to say go, but actually said gold.

Pair 5 (2) C3				
114	5H	Okay. Uh. **Gold** [arou], oh– Around the top …	Okay. **Go** around the top …	$O_{dis}=1$ $S=1$ $O_{dis}=2$

Example of **unintelligible**$_{disfluency}$ being counted as such.

Pair 5 (2) C3				
224	5H	Till you [unintelligible$_{disfluency}$] run into south bay.	Until you run into south bay.	$S=1$ $O_{dis}=1$

Example of an **interruption that is not an overlap**. Coded as O_{dis}.

Pair 5 (4) CF1				
209	5H	The old temple		$O_{turn}=1$ $O_{dis}=3$
	5D	Oh, wait, yes, I see the old temple. Okay, I've got that	Oh, wait, yes, I see the old temple. Okay, I've got that.	
216	5H	Go to the old temple.	Go to the old temple.	

Multiple word utterance transcribed into one word that has absolutely nothing to do with the utterance.

Pair 2 (6) CF1				
376	2H	Yeah.> below and then around baboons.	Clockwise.	$S=6$

Appendix 16: ASR-Dictation Scoring Rules

Compare ASR output with original text. Goal: Count and classify ASR transcription errors. Mark any differences on ASR output:

1. **Substitution** (red circle). The ASR transcribed a different word than the target.
    a. ASR collapsed 2 words into 1 word (e.g. 'of hearing' transcribed as 'appearing'), score as 2 errors.
    b. ASR transcribed on 1 word as 2 words (e.g. 'of the' transcribed as 'and'), score as 1 error.
2. **Morpheme Addition** (blue circle). ASR transcribed worded correctly but added a morpheme like 's' or 'ed'. Score as 1 error.
3. **Word/Morpheme Subtraction** (blue square). Score as 1 error.
4. **Hesitation** (green square). Talker, not ASR, error. Like a hesitation or a dysfluency. Do not count as an error. Note: Sometimes we were unsure of when to use this classification because we did not make audio recordings of these dictation sessions. Our general rule was to classify as a hesitation anything that didn't fit phonetically (in terms of sound) into the dictation.
5. **Ignore** (parentheses). Talker attempted commands to ASR, like 'new paragraph' or ' go to sleep', but ASR transcribed as text. Do not count as an error.
6. Tally all errors

Appendix 17: Conversational Games Analysis Manual

1. a. If the person says yes/no and continues to explain something further in the same utterance, code that as reply-y/n plus an explain.

adapted from Pr 4 (2) CF3					
	4H	3	query-yn		How about a pine tree?
43	4D		**reply-y**	3	Um, yeah.
	4D	3.1	**explain**		If you go from start and angle to the right, there's a sandstone cliffs and then there's a pine tree.

adapted from Pr 4 AF2, near 62							
	4H	10	query-yn		do you have a Stonecreek to the left of the rocks?	to have a Stonecreek to the left of the rocks	repetition
319	4D		**reply-n**	10	No, I don't. Um.		
	4D		**explain**	10	My Stonecreek is way down in the lower left-hand corner.		

b. When a person asks a question like "**<u>Do you have anything…</u>**," code it as **query-yn** (*even though the person's really asking, "What do you have?"*)
The response, then, can be [reply-y/n plus an explain] or [reply-w].

Pr 4 AF2, near 62						
	4H	3.1	query-yn		Do you have anything to the right of the diamond mine?	to you have anything to the right of the diamond mine
62	4D		**reply-y**	3.1	Yeah.	
	4D		**explain**	3.1	I have a cactus and death valley, and they're all the way to the right of the page.	
Pr 4 AF2, near 297						
	4H	9.4	query-yn		Do you have anything beneath the valley?	you have anything beneath the Valley
297	4D		**reply-w**	9.4	The [kak] cactus.	
adapted from Pr 4 A5, 59						
59	4H	3	query-yn		Do you have anything in the bottom left corner?	Give anything in the bottom left corner
	4D		**reply-y**	3	Yeah,	
	4D		**explain**	3	I have, the bottom left-hand corner I have a banana tree	
adapted from Pr 4 A5, 215						
215	4H	11	query-yn		Do you have hidden steps?	do have hidden steps
227	4D		**reply-y**	11	(unintelligible) Yeah, I have hidden steps. <Um/	
231	4D		**explain**	11	Above the waterfall, directly above the waterfall almost, um, almost to the top of the page.	
251	4H	11.2	query-yn		Near the waterfall about two inches up and one inch to the left?	nearer the waterfall about its two inches up in one inch to the left
310	4D		**reply-n**	11.2	No. No.>	

	4D		**explain**	11.2	The, the, the, the steps are, uh, I'm directly above the <waterfall/ But, um, quite a <ways./ Like, I don't know, seven or eight inches.	

2. Implicit information/messages/alternative suggestions
a. Coding should reflect intention of utterance, but **<u>DO NOT ADD MOVES</u>** that are not verbalized. (5/14/01). Do NOT do the following:

Pr 4 C1					
	4H	31	align		Are you now at the bottom left corner of the pyramid?
410	4D		**reply-n**	31	Bottom right corner.
	4D		**reply-w**	31	

Instead, do this:

Pr 4 C1					
	4H	31	align		Are you now at the bottom left corner of the pyramid?
410	4D		**reply-w**	31	Bottom right corner.

change to reply-w in the document itself as well as on this.
b. This rule also applies to the following situation: When a person suggests an **<u>alternative</u>** when asked about a specific landmark. (5/14/01)

adapted from Pr 4 A5, near 717 see pr 4 AF2, near 332						
	4H	25	query-yn		Do you have ancient ruins?	do have ancient ruins
724	4D		acknow	25	Ancient ruins,> ancient ruins, ancient ruins.	
			reply-w		Ruined city.	
adapted from Pr 4 A5, 787						
787	4H	27	query-yn		Do you have stones above> that?	Do you have stones ~~of love that~~
805	4D		acknow	27	Do you have [sto]–	
	4D		**reply-w**		I have fall [f□] uh, fallen rocks.	
adapted from Pr 4 AF2, near 332						
	4H	11	query-yn		Do you have Whitewater?	you have Whitewater
	4D		acknow		Whitewater. <Whitewater./	
	4D		**reply-w**		Rapids? I have rapids.	
adapted from Pr 4 AF2, just after 670						
(670)	4H	25	query-yn		Do you have a swan pond?	Do have a salon pond
	4D		**reply-n**	25	No,	
	4D	25.1	**explain**		I have a saloon.	
Pr 6 C2, near 102						
	6H	4	query-yn		Do you have a wagon wheel next <to the diamond mine?/	
107	6D	4.1	**query-w (aborted)**		So, how–> How about if I go to a swamp?	
Pr 6 F4, btn lines 54-68						
	6H	5.1	query-yn		Do you have a picnic site? (6H hand motion) <Picnic site/	Picnic site
68	6D	5.2	query-yn		<Would we settle for adventure playgrounds?/	
adapted from Pr 6 F4, near line 341						
	6H	18	query-yn		And do you have limestone cliffs?	
	6D		**reply-w**	18	I have lion country.	
	6D	18.3	**query-yn**		Could we settle for that?	
Pr 6 CF1, 161						

161	6H	7	query-yn		Do you have an old temple to the right of your paper above the polluted stream?
171	6D		acknow	7	Do I have something temple.
	6D		**reply-w**	7	I have a pyramid at the very top.

3. **Uncodeable Utterances**

a. Categories for moves aren't exhaustive – They don't cover everything. Some utterances are not easily classifiable as one of the possible moves listed. It's ok to not classify an utterance, put in ***no code***.

Pr 4 AF2, line 1						
1	4H	-1	ready		Okay.	~~Okay~~
	4H		*no code*		So your [s] start is at the top left.	~~so vigorous start is at the top lathes~~
	4H	1	query-yn		Um. Do you have a diamond mine?	~~onto the diamond mines~~

Put in the example with the crying in the bucket.

b. Insufficient information available from data to code – put in ***no code***.

Pr 4 F4, games 24-25 No AVI = No video recording						
	4H	24	query-yn		Are there limestone cliffs?	
	4D		*no code*		Limestone cliffs, limestone cliffs?	could be an **acknow** or check – not sure – no avi for this session
	4D		reply-n	24	No.	
	4H	25	query-yn		Is there an east lake?	
	4D		*no code*		East lake?	could be an **acknow** or check – not sure – no avi for this session
	4D		reply-y	25	Yeah.	

4. **Continued**

a. If you have an Instruct (Cont.) followed by an Acknow, don't mark that Acknow as Acknow (cont.).

b. When the first instruct plus cont. makes a complete sentence or though (could be a compound/complex sentence).

Pr 1 (4) A3						
	1H		reply-w	4.4	The line goes under the blacksmith.	The line goes under the blacksmith
187	1D		acknow	4.4	Uh-huh.	
188	1H		**reply-w (cont.)**		And curves up to the right, over an old pine tree.	in curves up to the right to over an old pine tree
	1H		asr		New paragraph. New paragraph.	~~depend~~

Pr 1 (4) A3						
	1D	12	query-w		Now what?	
376	1H		reply-w	12	The line goes to the canal.	bow line goes to the canal
382	1D		**acknow**	12	Uh-huh.	
	1H		**reply-w (cont.)**		Curves a little bit down.	curves a little bit down
	1D		**acknow**	13	Uh-huh.	
	1H	14	**reply-w (cont.)**		And then goes straight down	and then goes straight down three inches

					three inches.	
Pr 1 (4) A3 THIS ONE IS DIFFERENT BECAUSE IT ISN'T INSTRUCT-CONTINUED.						
	ID	19	query-w		Now what do I do?	
477	IH		reply-w	19	Curve under the wheatfield.	curve under the wheatfields
	ID		acknow		Uh-huh. Curve under…	
	IH	20	instruct		Slope to the left of the factory warehouse.	slope to the left of the factory warehouse
	ID		acknow	20	Factory warehouse.	
494	IH	21	instruct		Curve out to the left towards the edge of the paper.	curve out to the left towards the edge of the paper
	ID		acknow	21	Curve out towards the edge, okay. I'm at the edge of the paper now.	
	IH	22	instruct		Around to the old lighthouse.	around to the old Lighthouse
509	ID		acknow	22	Around to the old lighthouse.	

5. **Repetition**
[For future analysis] – Not using this yet.

Pr 4 F4, games 10.3						
	4H	**10**	**query-yn**		**Do you have a railroad crossing?**	
	4D		reply-y	10	Yeah,	
	4D	10.1	explain		the upper left-hand corner.	
	4H	10.2	query-yn		Is there one near the quarry?	
	4D	10.3	check		One in the other corner?	
	4H		reply-n	10.3	No.	
	4H		clarify	10.3	Near the granite quarry.	
	4D		acknow	10.2	Oh, yeah, okay. That's where I'm at.	**Breakdown**
	4H		query-yn	10.3	Is there a railroad crossing there?	**Repetition** - unelicitied
Pr 4 F4, games 10.3						
	4D		*no code*		Bottom edge of the lake?	could be an acknow or check – not sure – no avi
	4D		acknow	28.4	Okay, that's where I am.	
	4H		*no code*		So --	
	4D		acknow		The bottom shore east lake is where I am.	**Repetition**
adapted from Pr 4 A5, 450						
450	4D		reply-w	15.2	Go up three inches and go an inch and a half to the left. You have flamingos. They're, um, they're just below and to the left of the fallen pillars.	
461	4H	15.3	check		So between the bridge and the pillars?	So between the bridge and the pillars
476	4D		reply-y	15.3	Between the bridge and the pillars, the flamingos are, yeah, between the bridge and the pillars but to the left.	repetition of 15.2

6. Acknow – also used to code for **repetition of previous utterance** before answering a question.

adapted from Pr 4 A5, line 215						
215	4H	11	query-yn		Do you have hidden steps?	do have hidden steps
221	4D		**acknow**	11	Hidden steps, hidden steps, hidden steps. <Um/	
227	4D		reply-y	11	(unintelligible) Yeah, I have hidden steps. <Um/	
adapted from Pr 4 A5, near 717 see pr 4 AF2, near 332						
	4H	25	query-yn		Do you have ancient ruins?	do have ancient ruins
724	4D		**acknow**	25	Ancient ruins,> ancient ruins, ancient ruins.	
adapted from Pr 4 AF2, near 332						
	4H	11	query-yn		Do you have Whitewater?	you have Whitewater
	4D		**acknow**		Whitewater. <Whitewater./	

7. If an explain is in response to something the other person said, put the game number on the response (or *end*) column to indicate that some sort of exchange took place.

8. **READY & ACKNOW**

a. Don't have these two together for one conversational turn.
b. If you come across a ready for D, just code it as an acknow.
c. Don't code "Now" as READY – ignore for now.

9. **Request for Clarification**

a. [For future analysis] – Possible idea for a new question category.

Pr 4 A5, line 269						
269	4D	11.3	check		You're going> to have to repeat that.	**RQ**

b. Checks are for confirmation or to ask for repetition of the previous message.

adapted from Pr 4 (4) A5							
	4D		reply-w	25.1	That is above the fallen pillars about four inches and slightly to the left.		
743	4H		**check**		Say that again?	say that again	**RQ**

10. **Unusual Utterances**

adapted from Pr 1 (4) A3						
	1D		explain	2.6	my X is kind of above the cliff.	
	1D	2.7	**query-w**		so I need to know if I'm going over the cliffs or under the cliffs.	
	1H		reply-w	2.7	Next to the cliffs.	next to the Cliffs
Pr 1 (2) AF2, right after 346						
346	1D	2.3	query-w		Okay. So tell me more about the swan pond.	

Appendix 18: Conversational Games Analysis Summary

Pair 1	F	CF	C	AF	A
EXPLAIN	17	12	13	15	13
INSTRUCT	19	22	15	8	10
QUERY	10	11	13	33	19
ALIGN	0	1	1	0	1
CHECK	20	16	9	6	7
ACKN	50	49	30	33	36
CLARIFY	4	7	2	3	2
REPLY	27	23	21	31	22
TOTAL	147	141	104	129	110
ASR MGMT				1	13
Percent	F	CF	C	AF	A
EXPLAIN	12%	9%	13%	12%	12%
INSTRUCT	13%	16%	14%	6%	9%
QUERY	7%	8%	13%	26%	17%
ALIGN	0%	1%	1%	0%	1%
CHECK	14%	11%	9%	5%	6%
ACKN	34%	35%	29%	26%	33%
CLARIFY	3%	5%	2%	2%	2%
REPLY	18%	16%	20%	24%	20%
sum	100%	100%	100%	100%	100%

Pair 2	F	CF	C	AF	A
EXPLAIN	12	34	26	21	18
INSTRUCT	26	24	27	32	25
QUERY	32	6	10	15	21
ALIGN	11	9	8	11	7
CHECK	36	14	11	18	7
ACKN	57	31	49	53	54
CLARIFY	20	10	4	8	2
REPLY	41	27	22	45	37
TOTAL	235	155	157	203	171
ASR MGMT				42	72
Percent	F	CF	C	AF	A
EXPLAIN	5%	22%	17%	10%	11%
INSTRUCT	11%	15%	17%	16%	15%
QUERY	14%	4%	6%	7%	12%
ALIGN	5%	6%	5%	5%	4%
CHECK	15%	9%	7%	9%	4%
ACKN	24%	20%	31%	26%	32%
CLARIFY	9%	6%	3%	4%	1%
REPLY	17%	17%	14%	22%	22%
sum	100%	100%	100%	100%	100%

Pair 3	F	CF	C	AF	A
EXPLAIN	5	9	6	4	5
INSTRUCT	20	30	29	23	17
QUERY	23	19	19	18	15
ALIGN	11	6	4	9	8
CHECK	19	9	9	4	7
ACKN	22	30	29	27	18
CLARIFY	4	4	6	2	5
REPLY	39	27	27	25	22
TOTAL	143	134	129	112	97
ASR MGMT				0	49
Percent	F	CF	C	AF	A
EXPLAIN	3%	7%	5%	4%	5%
INSTRUCT	14%	22%	22%	21%	18%
QUERY	16%	14%	15%	16%	15%
ALIGN	8%	4%	3%	8%	8%
CHECK	13%	7%	7%	4%	7%
ACKN	15%	22%	22%	24%	19%
CLARIFY	3%	3%	5%	2%	5%
REPLY	27%	20%	21%	22%	23%
sum	100%	100%	100%	100%	100%

Pair 4	F	CF	C	AF	A
EXPLAIN	7	31	1	12	13
INSTRUCT	18	17	23	14	11
QUERY	14	18	15	20	17
ALIGN	2	7	2	13	5
CHECK	11	9	5	15	7
ACKN	29	60	35	46	34
CLARIFY	1	1	1	3	0
REPLY	24	32	22	40	25
TOTAL	106	175	104	163	112
ASR MGMT				54	58
Percent	F	CF	C	AF	A
EXPLAIN	7%	18%	1%	7%	12%
INSTRUCT	17%	10%	22%	9%	10%
QUERY	13%	10%	14%	12%	15%
ALIGN	2%	4%	2%	8%	4%
CHECK	10%	5%	5%	9%	6%
ACKN	27%	34%	34%	28%	30%
CLARIFY	1%	1%	1%	2%	0%
REPLY	23%	18%	21%	25%	22%
sum	100%	100%	100%	100%	100%

Pair 5	F	CF	C	AF	A
EXPLAIN	19	26	7	15	18
INSTRUCT	24	23	16	23	21
QUERY	6	13	3	9	9
ALIGN	1	3	4	5	4
CHECK	20	3	1	7	15
ACKN	16	24	20	18	25
CLARIFY	10	0	0	0	5
REPLY	12	12	7	19	25
TOTAL	108	104	58	96	122
ASR MGMT				3	0
Percent	F	CF	C	AF	A
EXPLAIN	18%	25%	12%	16%	15%
INSTRUCT	22%	22%	28%	24%	17%
QUERY	6%	13%	5%	9%	7%
ALIGN	1%	3%	7%	5%	3%
CHECK	19%	3%	2%	7%	12%
ACKN	15%	23%	34%	19%	20%
CLARIFY	9%	0%	0%	0%	4%
REPLY	11%	12%	12%	20%	20%
sum	100%	100%	100%	100%	100%

Pair 6	F	CF	C	AF	A
EXPLAIN	12	7	12	9	12
INSTRUCT	17	13	11	11	14
QUERY	13	16	15	7	13
ALIGN	0	0	0	1	0
CHECK	9	9	2	8	18
ACKN	37	36	27	21	27
CLARIFY	0	0	0	0	1
REPLY	17	24	16	15	13
TOTAL	105	105	83	72	98
ASR MGMT				0	0
Percent	F	CF	C	AF	A
EXPLAIN	11%	7%	14%	13%	12%
INSTRUCT	16%	12%	13%	15%	14%
QUERY	12%	15%	18%	10%	13%
ALIGN	0%	0%	0%	1%	0%
CHECK	9%	9%	2%	11%	18%
ACKN	35%	34%	33%	29%	28%
CLARIFY	0%	0%	0%	0%	1%
REPLY	16%	23%	19%	21%	13%
sum	100%	100%	100%	100%	100%

Summary of Totals

Percent					
All Pairs	F	CF	C	AF	A
EXPLAIN	9%	14%	10%	10%	11%
INSTRUCT	16%	16%	20%	15%	14%
QUERY	11%	11%	12%	13%	13%
ALIGN	3%	3%	3%	5%	3%
CHECK	13%	7%	5%	7%	9%
ACKN/READY	25%	28%	31%	25%	27%
CLARIFY	4%	2%	2%	2%	2%
REPLY	19%	18%	18%	22%	20%
sum	100%	100%	100%	100%	100%

Appendix 19: Raw Data

										Partic. Eval	
Pair	Speech-reading	Hearing Loss	Cond	Map ID	Test Order	Trans. Acc.	Map Score	Time (min)	Check Moves	Deaf	Hear
1	95%	prof	F	1	3		64%	7.2	0.14	2	1
			CF	5	5	92%	64%	6.4	0.11	1	2
			C	4	6	98%	92%	6.6	0.09	4	3
			AF	2	2	85%	64%	6.6	0.05	3	4
			A	3	4	90%	69%	9.6	0.06	5	5
2	97%	sev	F	2	2		64%	10.5	0.15	1	2
			CF	1	6	86%	93%	8.0	0.09	2	1
			C	4	5	92%	77%	12.0	0.07	4	3
			AF	5	4	73%	64%	13.5	0.09	4	4
			A	3	3	89%	77%	12.8	0.04	4	5
3	75%	prof	F	4	4		62%	6.5	0.13	2	3
			CF	3	3	98%	69%	12.5	0.07	1	1
			C	5	5	98%	50%	13.8	0.07	1	2
			AF	2	6	75%	50%	10.0	0.04	3	3
			A	1	2	87%	43%	15.0	0.07	5	5
4	93%	sev	F	4	5		85%	9.0	0.10	3	3
			CF	3	2	96%	85%	11.5	0.05	1	1
			C	1	6	100%	93%	7.4	0.05	2	2
			AF	2	3	81%	43%	15.0	0.09	4	4
			A	5	4	92%	64%	15.0	0.06	5	5
5	44%	prof	F	5	6		29%	8.7	0.19	5	3
			CF	1	4	95%	36%	8.9	0.03	1	1
			C	3	2	92%	31%	5.6	0.02	3	2
			AF	4	5	76%	62%	15.5	0.07	2	4
			A	2	3	76%	21%	13.9	0.12	4	5
6	45%	prof	F	4	2		69%	6.4	0.09	3	4
			CF	1	5	98%	43%	8.7	0.09	2	1
			C	2	4	99%	43%	9.1	0.02	1	2
			AF	5	3	75%	50%	8.5	0.11	4	3
			A	3	6	86%	31%	13.0	0.18	5	5

Appendix 20: Check Moves Post-hoc analyses

1-way within subjects ANOVA (repeated measures)

CONDITION	n	Mean	SD	SE	
FTF	6	0.236	0.033	0.0136	
CART-FTF	6	0.171	0.039	0.0160	
Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.013	1	0.013	7.05	**0.0451**
Between CONDITION	0.004	5	0.001	0.46	0.7901
Within cells	0.009	5	0.002		
Total	0.026	11			

CONDITION	n	Mean	SD	SE	
FTF	6	0.236	0.033	0.0136	
CART	6	0.142	0.043	0.0174	
Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.026	1	0.026	18.10	**0.0081**
Between CONDITION	0.007	5	0.001	1.03	0.4857
Within cells	0.007	5	0.001		
Total	0.041	11			

CONDITION	n	Mean	SD	SE	
FTF	6	0.236	0.033	0.0136	
ASR-FTF	6	0.173	0.037	0.0149	
Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.012	1	0.012	6.88	**0.0470**
Between CONDITION	0.004	5	0.001	0.43	0.8130
Within cells	0.009	5	0.002		
Total	0.024	11			

CONDITION	n	Mean	SD	SE	
FTF	6	0.236	0.033	0.0136	
ASR	6	0.189	0.056	0.0227	
Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.006	1	0.006	2.39	**0.1829**
Between CONDITION	0.007	5	0.001	0.55	0.7379
Within cells	0.014	5	0.003		
Total	0.028	11			

Appendix 20, continued **1-way within subjects ANOVA (repeated measures)**

CONDITION	n	Mean	SD	SE
ASR-FTF	6	0.074	0.029	0.0117
ASR	6	0.091	0.053	0.0216

Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.001	1	0.001	0.75	**0.4269**
Between CONDITION	0.013	5	0.003	2.33	0.1873
Within cells	0.005	5	0.001		
Total	0.019	11			

CONDITION	n	Mean	SD	SE
CART-FTF	6	0.171	0.039	0.0160
ASR	6	0.189	0.056	0.0227

Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.001	1	0.001	0.37	**0.5681**
Between CONDITION	0.009	5	0.002	0.63	0.6877
Within cells	0.014	5	0.003		
Total	0.024	11			

CONDITION	n	Mean	SD	SE
CART	6	0.142	0.043	0.0174
ASR	6	0.189	0.056	0.0227

Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.007	1	0.007	1.51	**0.2744**
Between CONDITION	0.003	5	0.001	0.12	0.9816
Within cells	0.022	5	0.004		
Total	0.031	11			

CONDITION	n	Mean	SD	SE
CART-FTF	6	0.171	0.039	0.0160
ASR-FTF	6	0.173	0.037	0.0149

Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.000	1	0.000	0.01	**0.9175**
Between CONDITION	0.006	5	0.001	0.79	0.5998
Within cells	0.008	5	0.002		
Total	0.014	11			

CONDITION	n	Mean	SD	SE
CART	6	0.142	0.043	0.0174
ASR-FTF	6	0.173	0.037	0.0149

Source of variation	SSq	DF	MSq	F	p
Within CONDITION	0.003	1	0.003	1.13	**0.3357**
Between CONDITION	0.003	5	0.001	0.27	**0.9096**
Within cells	0.012	5	0.002		
Total	0.019	11			

Appendix 21: Participant Evaluation Post-hoc analyses

Wilcoxon signed ranks test

Participant Evaluation: CART-FTF ≠ CART-only

n	12		
Difference between pairs	**n**	**Rank sum**	**Mean rank**
Positive	1	4.0	4.00
Negative	10	62.0	6.20
Zero	1		
Difference between medians	-1.000		
95.8% CI	-2.000 to -0.500	(exact)	
Wilcoxon's W statistic	4		
2-tailed p	0.0068 (exact tables used, 91% ties)		

Participant Evaluation: ASR-only ≠ ASR-FTF

n	12		
Difference between pairs	**n**	**Rank sum**	**Mean rank**
Positive	11	66.0	6.00
Negative	0	0.0	-
Zero	1		
Difference between medians	1.500		
95.8% CI	1.000 to 2.000	(exact)	
Wilcoxon's W statistic	66		
2-tailed p	0.0010 (exact tables used, 100% ties)		

Participant Evaluation: FTF-only ≠ ASR-FTF

Difference between pairs	**n**	**Rank sum**	**Mean rank**
Positive	2	14.0	7.00
Negative	9	52.0	5.78
Zero	1		
Difference between medians	-1.000		
95.8% CI	-2.000 to 0	(exact)	
Wilcoxon's W statistic	14		
2-tailed p	0.1016 (exact tables used, 91% ties)		

Appendix 21. Continued **Wilcoxon signed ranks test**

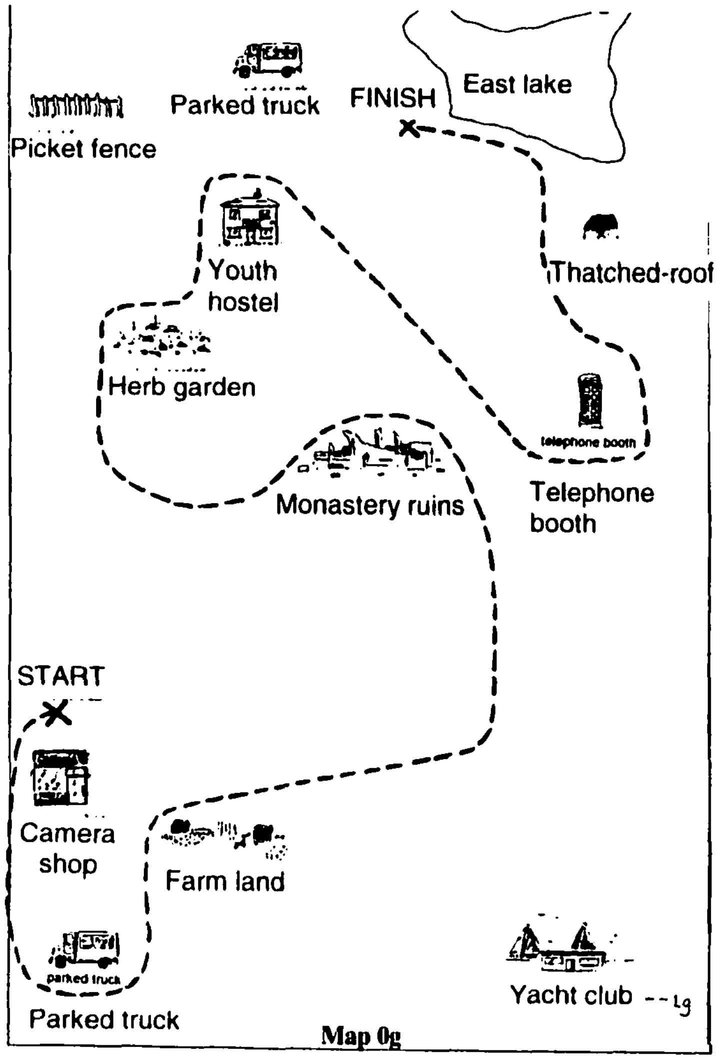
Participant Evaluation: FTF-only ≠ CART-only

n	12		
Difference between pairs	**n**	**Rank sum**	**Mean rank**
Positive	8	44.5	5.56
Negative	4	33.5	8.38
Zero	0		
Difference between medians	0.250		
95.8% CI	-1.000 to 1.500		(exact)
Wilcoxon's W statistic	44.5		
2-tailed p	0.7334 (exact tables used, 92% ties)		

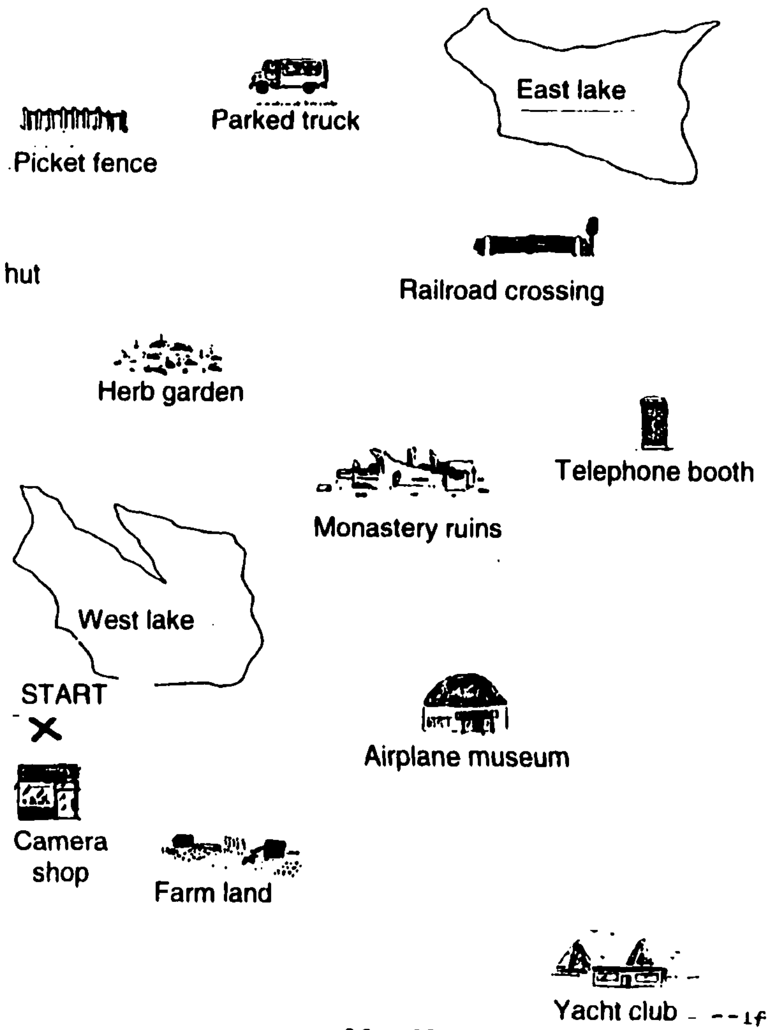
Participant Evaluation: CART-only ≠ ASR-FTF

n	12		
Difference between pairs	**n**	**Rank sum**	**Mean rank**
Positive	2	7.0	3.50
Negative	9	59.0	6.56
Zero	1		
Difference between medians	-1.000		
95.8% CI	-2.000 to 0		(exact)
Wilcoxon's W statistic	7		
2-tailed p	0.0186 (exact tables used, 91% ties)		

Appendix 22: The Map Task Stimuli

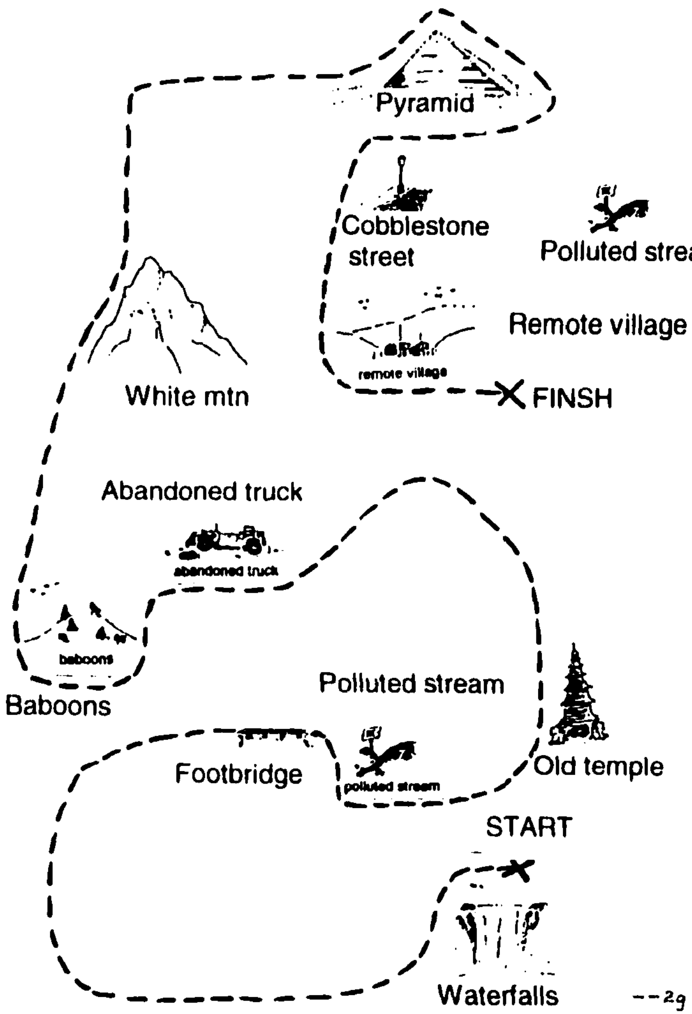


Map 0g

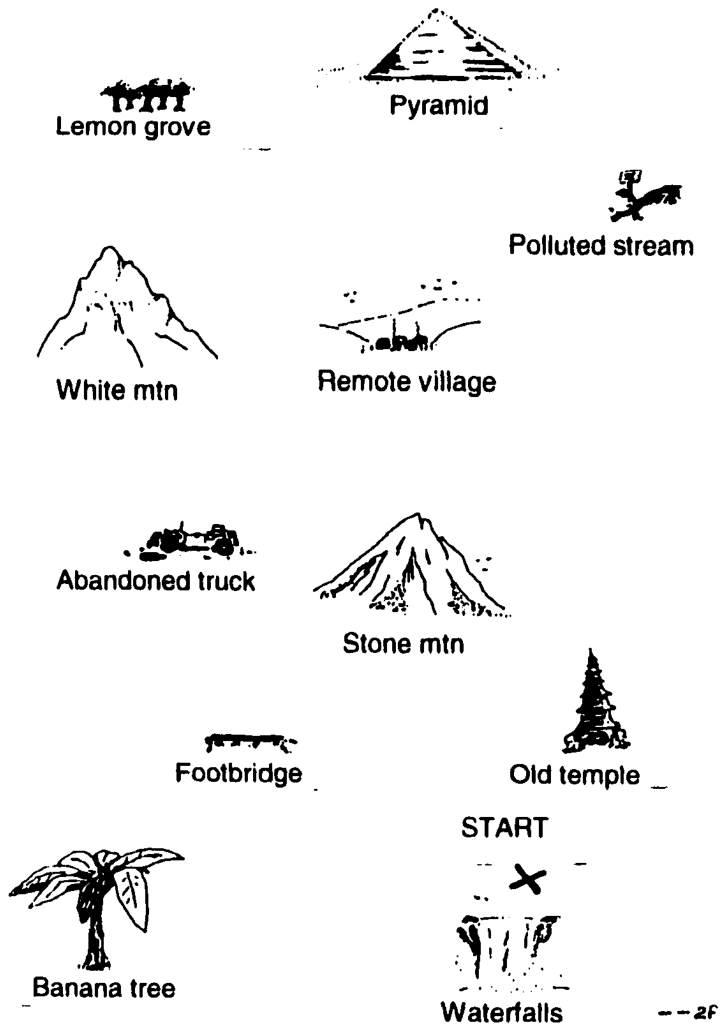


Map 0f

157

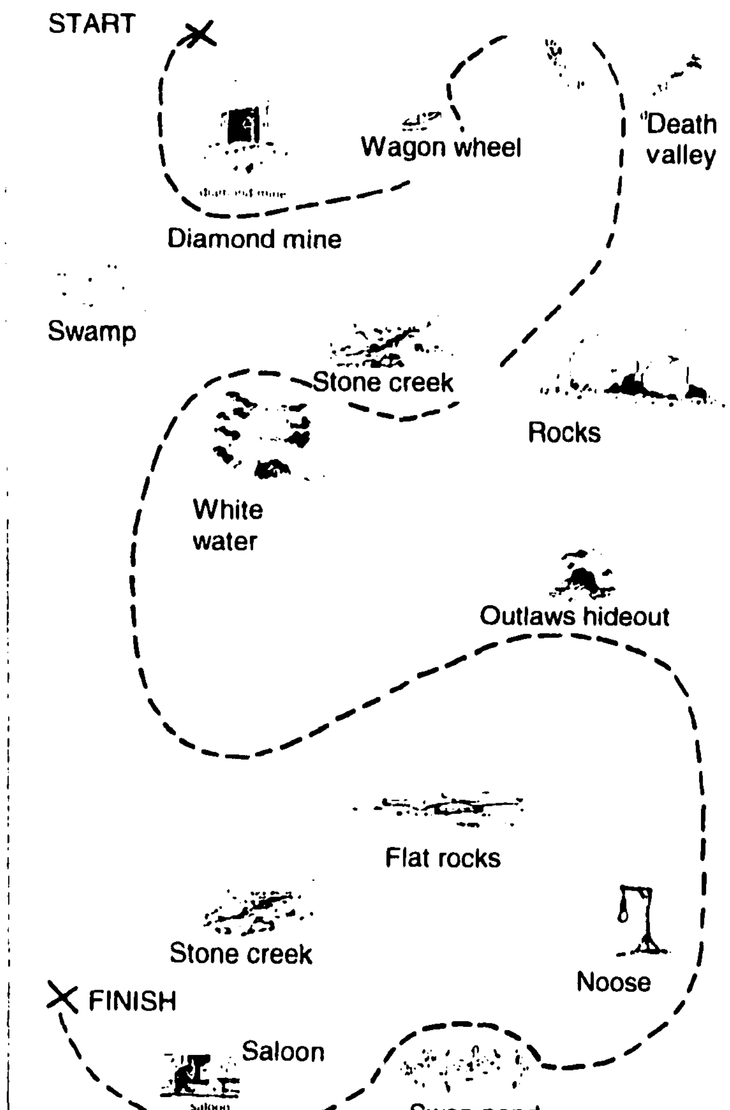


Map 1g

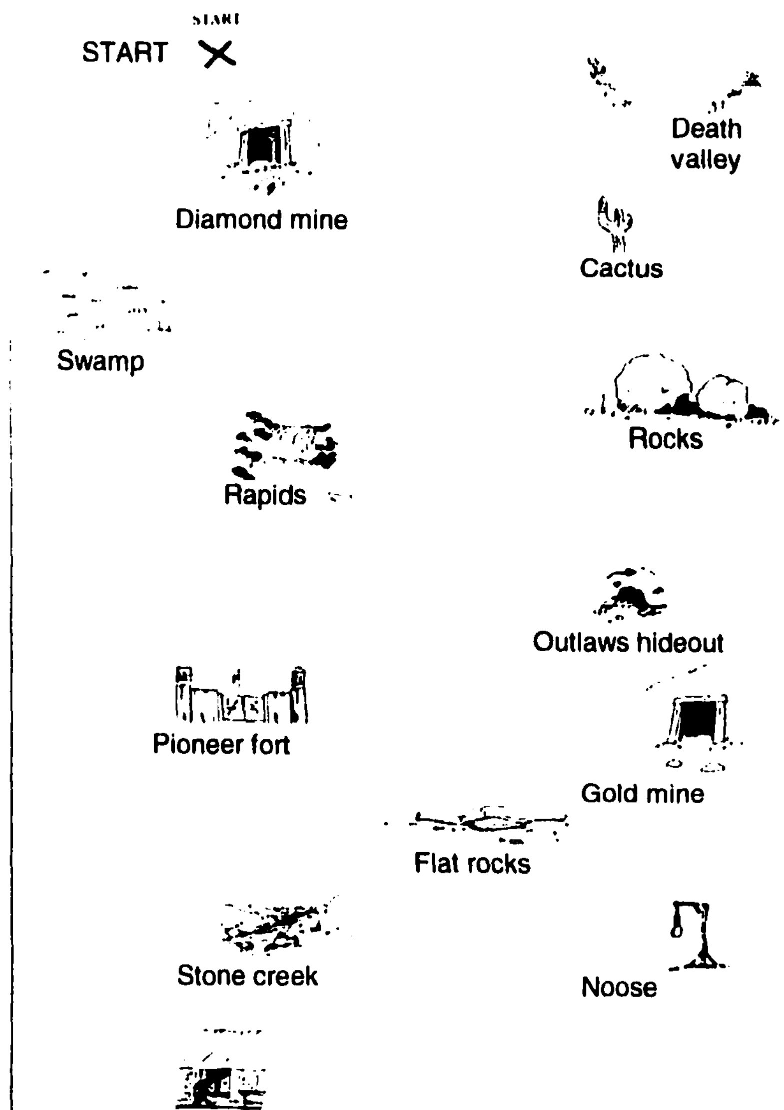


Map 1f

158

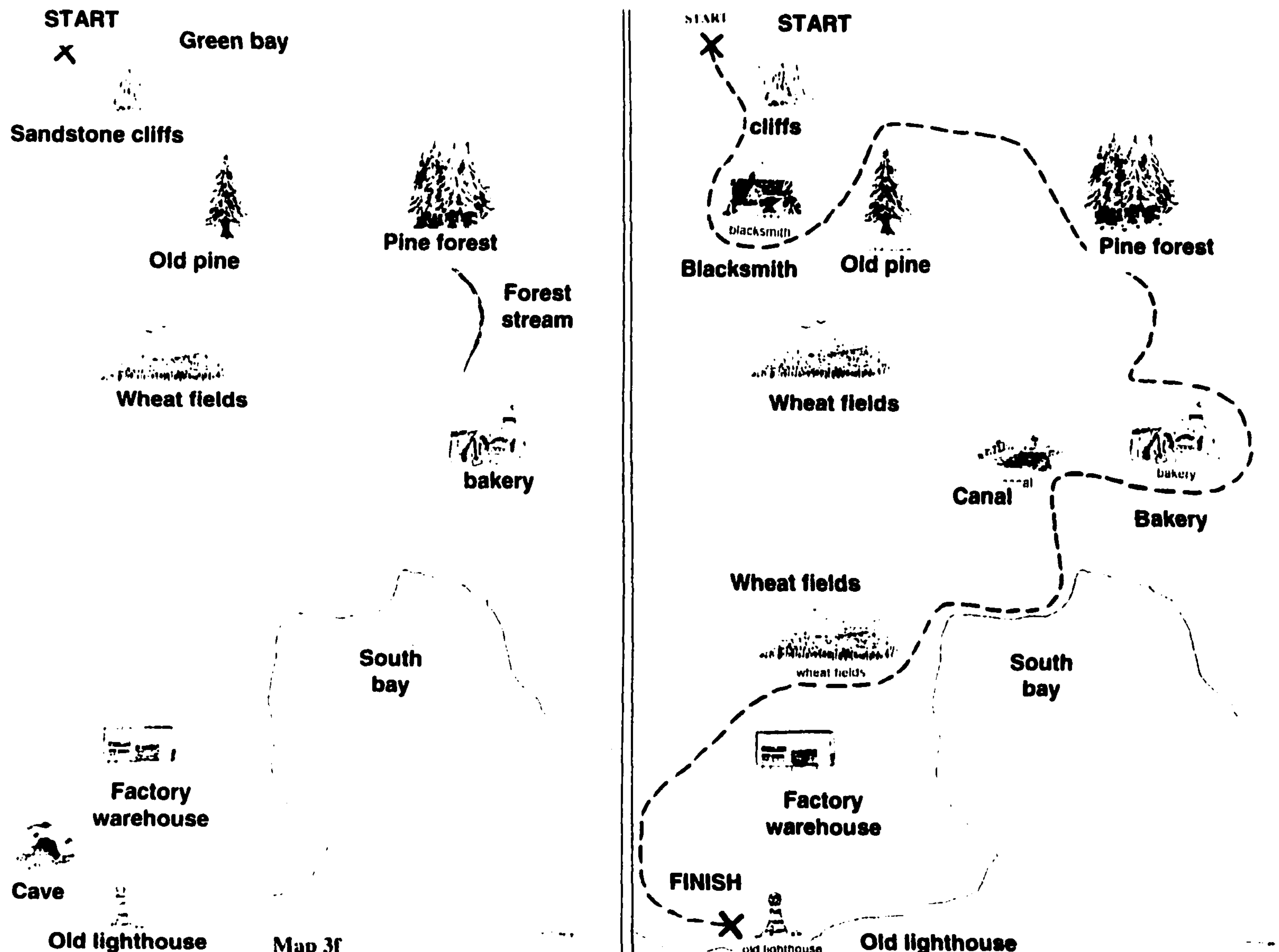


Map 2b

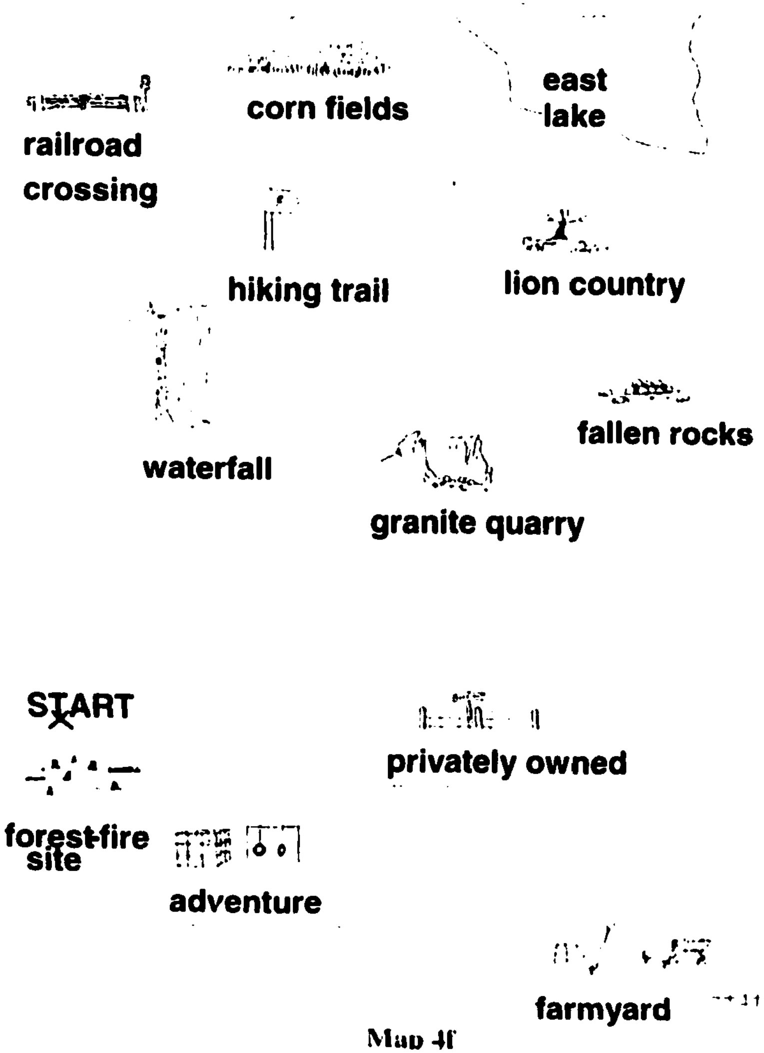


Map 2f

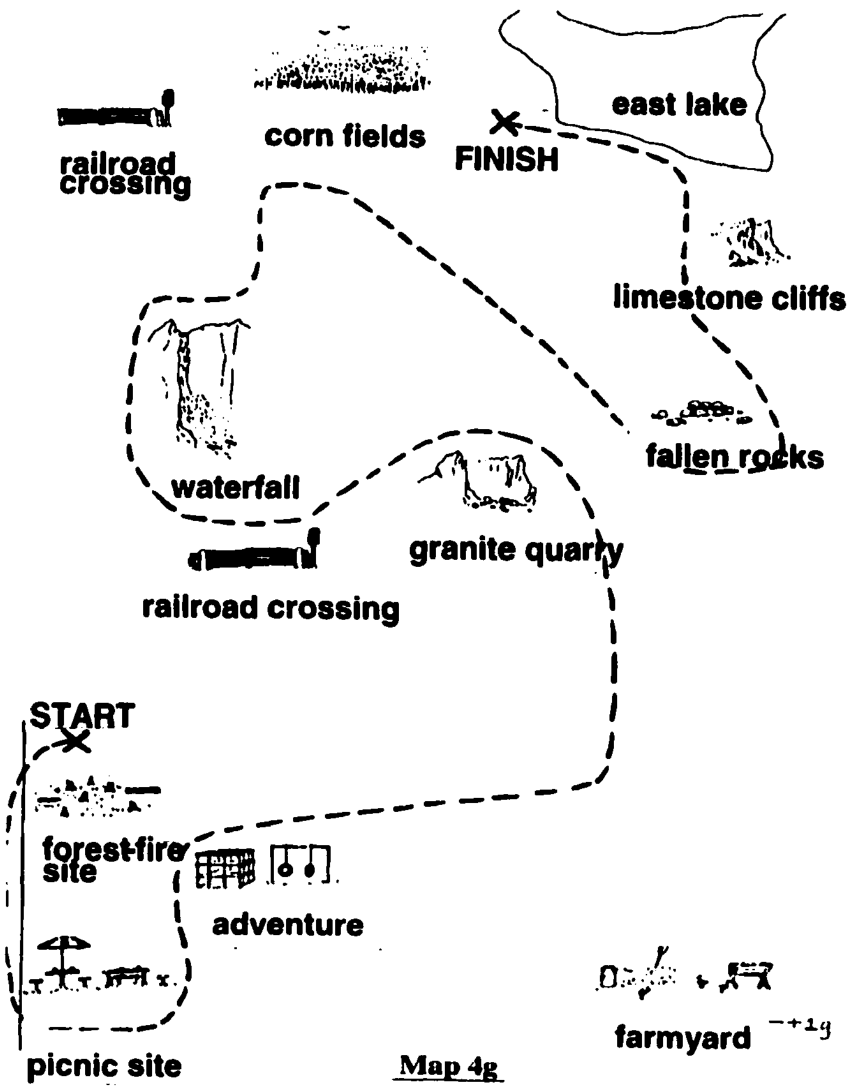
159


START
Green bay
Sandstone cliffs
Old pine
Pine forest
Forest stream
Wheat fields
bakery
South bay
Factory warehouse
Cave
Old lighthouse
START
cliffs
blacksmith
Blacksmith
Old pine
Pine forest
Wheat fields
Canal
bakery
Bakery
Wheat fields
wheat fields
South bay
Factory warehouse
FINISH
old lighthouse
Old lighthouse

Map 3f.

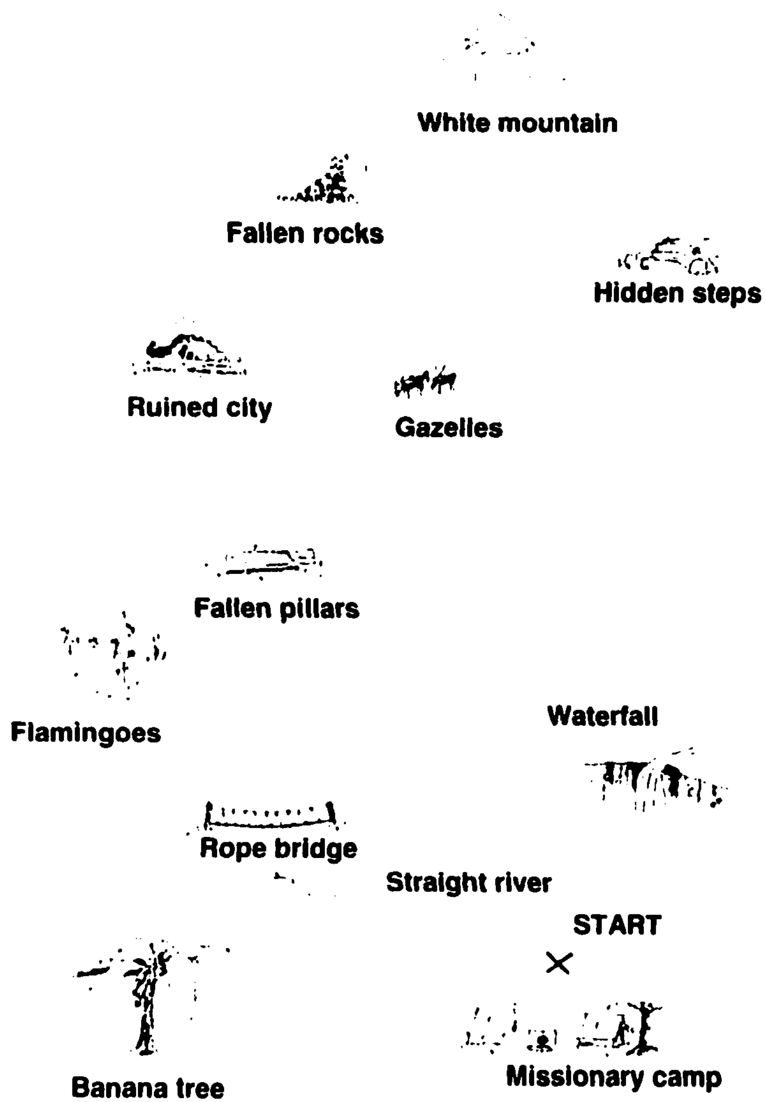


Map 4f

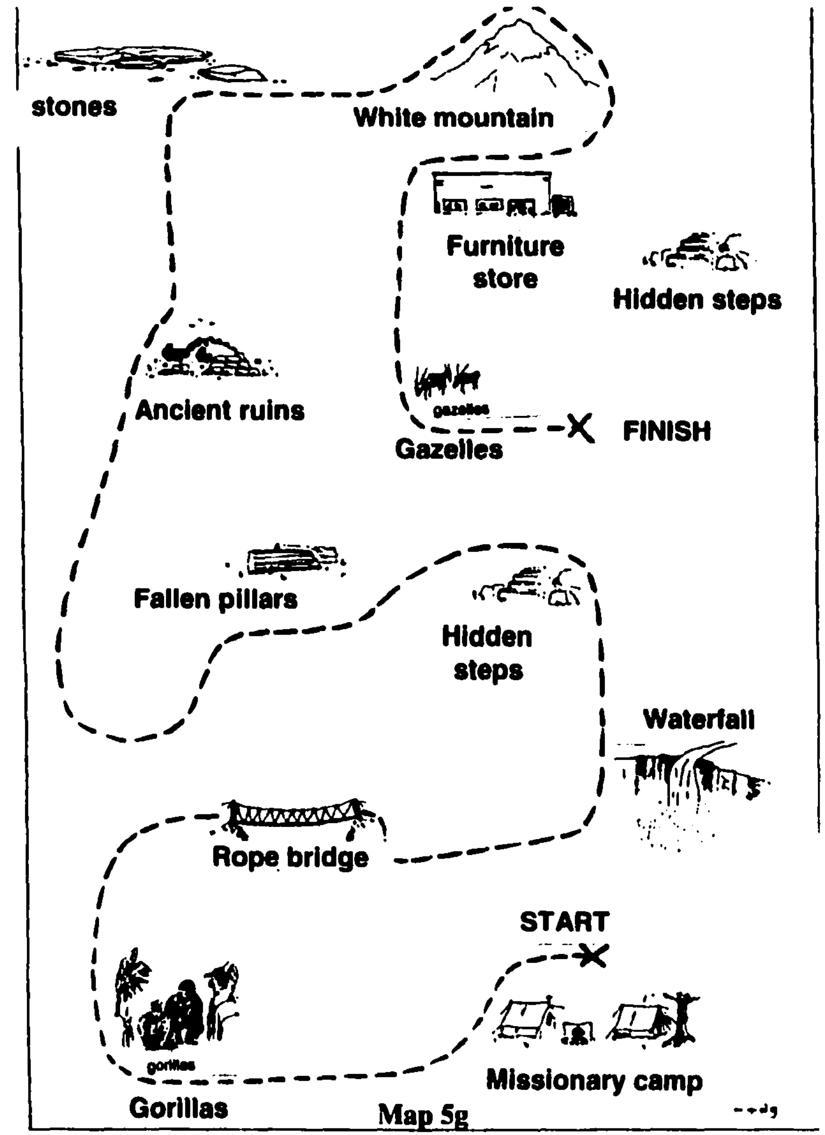


Map 4g

161



Map 5f



Map 5g

162

REFERENCES

Acker, S.R. & Levitt S.R. (1987). Designing videoconference facilities for improved eye contact. *Journal of Broadcasting & Electronic Media, 31*, 181-191.

Analyse-it statistics add-in for Microsoft Excel, vsn1.62 (2001). PO Box 77, Leeds, UK. http://www.analyse-it.com/

Anderson A.H. & Boyle, E.A. (1994). Forms of introduction in dialogues: their discourse contexts and communicative consequences. *Language and Cognitive Processes, 9*, 101-122.

Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko. J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech, 34*, 351-366.

Anderson, A.H., Clark, A., & Mullin, J. (1991). Introducing information in dialogues: forms of introduction chosen by young speakers and the responses elicited from young listeners. *Child Language, 18*, 663-687.

Anderson, A.H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A.M., & Van der Velder, J. (1997). The impact of VMC on collaborative problem solving: an analysis of task performance, communicative process and user satisfaction. In Finn, K.E., Sellen, A.J. & Wilbur, S.B., Eds. *Video-mediated Communication*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baig, E.C. (2000) Look, Ma: No hands, no ums, no ahs. *USA Today*: Gannet Co., Inc.

Bernstein, L.E., Demorest, M.E. & Tucker, P.E. (1998). What makes a good speechreader? First you have to find one. In Campbell, R, Dodd, B., & Burnham, D., Eds. *Hearing by eye II: advances in the psychology of speechreading and auditory-visual speech*. Hove: Psychology Press.

Boothroyd, A. (1998). The perception of speech by children. In Weisel, A., Ed. *Issues Unresolved: New Perspectives In Language And Deaf Education*. Washington, DC: Gallaudet University Press.

Boothroyd, A. (1988). Linguistic factors in speechreading. *Volta Review, 90(5)*, 77-87.

Boothroyd, A., Hnath-Chisolm, T. Hanin, L. & Kishon-Rabin, L. (1988). Voice fundamental frequency as an auditory supplement to the speechreading of sentences. *Ear and Hearing*, 9: 306-312.

Boyle, E.A., Anderson, A.H. & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37, 1-20.

Brennan, S.E. (2000). Processes that shape conversation and their implications for computational linguistics. *Proceedings, 38th Annual Meeting of the Association for Computational Linguistics*. Hong Kong: ACL.

Brown, G. Anderson, A.H., Yule, G., & Shillcock, R. (1984). *Teaching Talk*. Cambridge, England: Cambridge Univ. Press.

Caissie, R & Gibson, C.L.(1997). The effectiveness of repair strategies used by people with hearing losses and their conversational partners. *Volta Review, 99(4)*, 203-218.

Caissie, R. & Rockwell, E. (1993). A videotape analysis procedure for assessing conversational fluency in hearing-impaired adults. *Ear and Hearing*, 14(3), 202-209.

Caissie, R. Dawe, A.L., Donovan C., Brooks, H. & MacDonald, S.M. (1998). Conversational performance of adults with hearing loss. *Journal of the Academy of Rehabilitative Audiology*, 31, 31-44.

Castle, D.L. (1994). Telecommunications: visual technology. In Ross, M. (Ed.), *Communication Access for Persons with Hearing Loss*. Baltimore MD: York Press, 145-166.

Campbell, R, Dodd, B., & Burnham, D., Eds. (1998). *Hearing by eye II: advances in the psychology of speechreading and auditory-visual speech*. Hove: Psychology Press.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Association for Computational Linguistics, 22*, 249-254.

Carletta, J., Isard. A., Isard, S., Kowto, J.C., Doherty-Sneddon, G., & Anderson, A.H. (1997). The reliability of a dialogue structure coding scheme. *Association for Computational Linguistics*, 23, 13-31.

Clark, H.H. & Brennan, S.E. (1991). Grounding in communication. In Resnick, L.B., Levine, J.M. & Teasley, S.D. (Eds.), *Perspectives on Socially Shared Cognition*, Washington, D.C: American Psychological Association.

Clark, H.H. & Schaefer, F.S. (1989). Contributing to discourse. *Cognitive Science 13*, 259-294.

Clark, H.H. (1992). *Arenas of Language Use*. Chicago: University of Chicago Press.

Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

Coco, D. S. (2000). Speeding up relay services via ASR. *Hearing Health, Jan/Feb*, 82-84.

Coco, D. S. & Bagnall, J. (2000). The liberated learning project: improving access for persons with disabilities in higher education using automated speech recognition. PEPNet 2000 Proceedings.

Conover, W.J. (1980). Practical Non-parametric Statistics (2nd Edition). ISBN 0-471-02867-3; 122-128.

Cutler, A., & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. In C Johns-Lewis (Ed.) *Intonation in Discourse* (pp.139-155). London: Croom Helm.

De La Paz, S. (1999). Composing via dictation and speech recognition systems: compensatory technology for students with learning disabilities. *Learning Disability Quarterly, 22,* 173-182.

Doherty-Sneddon, G., Anderson, A. O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: a comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3,* 105-125.

Erber, N.P. (1988). *Communication therapy for hearing-impaired adults.* Abbotsford, Victoria, Australia: Clavis Publishing.

Fay, E.A. (1883). The glossograph. *American Annals of the Deaf and Dumb, 28,* 67-69.

Foster, S. (1992). *Working with Deaf People: Accessibility and Accommodations in the Workplace.* Springfield, IL: Charles C. Thomas.

Gagne, J.-P., Stelmacovich, P., & Yovetich, W. (1991). Reactions to requests for clarification used by hearing-impaired individuals. *The Volta Review*, 129-143.

Gibson, C., & Caissie, R. (1994). The effectiveness of repair strategy intervention with a hearing-impaired adult. *Journal of Speech Language Pathology, 18,* 14-21.

Grice, H.P. (1978). Some further notes on logic and conversation. In P. Cole (Ed.) *Syntax and semantics, Volume 9: Pragmatics* (pp. 113-128). New York. Academic Press.

Hallberg, L.R.-M. & Carlsson S.G. (1991). A qualitative study of strategies for managing a hearing impairment. *British Society of Audiology*, 201-211.

Haravon, A.B. (July, 2000). Grounding communication with deaf and hard-of-hearing people: technological advances. Paper presented at *the Seventh International Pragmatics Conference*, Budapest, Hungary.

Haravon, A.B. (March, 2000). Grounding communication with deaf and hard-of-hearing people: technological advances. Paper presented at the *First CUNY/SUNY Linguistics Mini-Conference*, New York, NY.

Haravon, A.B. (1999). Automatic speech recognition as a communication aid for people with hearing loss. Paper presented at the *Language and Communication: Breakdown and Theory Conference*, Hebrew University, Jerusalem, Israel.

Haravon, A.B. & Harkins, J.E. (1999). Automatic speech recognition as a communication aid for people with hearing loss: a hands-on workshop. Workshop conducted at the *Eleventh International Association of Late-Deafened Adults* Convention. Alexandria, VA.

Haravon, A.B. Harkins, J.E., Obler, L.K. & Eum, J. (November, 2001). Discourse and hearing loss: the effects of a text supplement. Paper presented at the *American Speech and Hearing Association 2001 Convention*, New Orleans, LA.

Haravon, A.B. Harkins, J.E., Obler, L.K. & Levitt, H. (March, 2002). Speech-to-text technologies for deaf people. Poster presented at the *American Auditory Society Conference*, Scottsdale, AZ.

Haravon, A.B., Obler, L.K., Sarno, M.T. (1994). A method for microanalysis of discourse in brain-damaged patients. In Bloom et al., *Discourse in Adult Clinical Populations*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Harkins, J.E. & Bakke M. (in press). Advances in Communication Technology. In M. Marschark & P. Spencer (Eds.) *Handbook of Deaf Studies, Language, and Education*. Oxford University Press.

Harkins, J.E., Haravon, A.B., Tucker, P.E., Eum, J. & Fisher, L. (June, 2001.). Evaluating automatic speech recognition as a conversational tool for people with hearing loss. Paper to be presented at the *Rehabilitation Engineering Society of North America 2001 Annual Conference*, Reno, NV.

Harkins, J. & Jensema, C.J. (1988). Focus-group discussions with deaf and hard of hearing people on needs for sensory devices. *Journal of the American Deafness and Rehabilitation Association.*

Higginbotham, D.J., Scally C.A., Lundy, D.C. & Kowarsky, K. (1995). Discourse comprehension of synthetic speech across three augmentative and alternative communication (AAC) output methods. *Journal of Speech and Hearing Research, 38*, 889-901.

Higginbotham, D.J. & Yoder, D.E. (1982). Communication within natural interaction: imlications for severe communicatively impaired persons. *Topics in Language Disorders*, March, 1982.

Ichikawa, Y., Okada K., Jeong, G., Tanaka, S. & Matsushita, Y. (1995). MAJIC Videoconferencing system: experiments, evaluation and improvement. *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work*, Electronic Meetings II, 279-292.

Ishii, H. & Kobayashi, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye chart. *Computer Human Interface*, 525-532.

Jefferson, G.A. (1973) A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica, 9*, 47-49.

Kowtko, JC, Isaard, S, & Doherty-Sneddon (1991, April). Conversational Games in dialogue. In A. Lescardes (Ed.), *Proceedings of the ESPRIT/DAWDI Workshop On Discourse Coherence.* Edinburgh, Scotland,: Univ. of Edinburgh.

Knowles, W.B. (1972). Human factors. *The Journal of Human Factors Society, 14*, 487-509.

Lamb, S.J. (1996) Supporting children's communication: activities used as part of the language and communication project. *Technical Report No. 38*. ESRC Centre for Research in Development, Instruction and Training, University of Nottingham.

Lehman, E.L. (1975). *Non-parametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day; 175-85.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levitt, H. (1995). Sensory processing for physical and sensory disabilities. *National Academy of Sciences Proceedings, 92*, 311-343.

Levitt, H. (1998). Automatic speech recognition and its applications. In A. Weisel (Ed.). *Issues Unresolved: New Perspectives on Language and Deaf Education.* Washington, D.C.: Gallaudet University Press.

Levy, E. & Nelson, K. (1994). Words in discourse: a dialectical approach to the acquisition of meaning and use. *Child Language, 18*, 367-389.

Makhoul J. & Schwartz, R. (1995). State of the art in continuous speech recognition. *National Academy of Sciences Proceedings*. 92, 165-198.

Murphy, J. (1999). Don't hang up: use of the telephone by people with communciation difficulties. *Nursing Standard, 13*, 33-36.

Murphy, J. (1998). AAC users' interaction with their peers. *Communciation Matters. 12 (1)*, 27-29.

Murphy, J. Markova, I, Collins, S. & Moodie, E. (1996). AAC Systems: obstacles to effective use. *European Journal of Disorders of Communication, 31*, 31-44.

Newlands, A. (1998) The effects of computer-mediated-communication on the process of communication and collaboration. Thesis (Ph.D.) University of Glasgow.

Newlands, A., Anderson, A.H. & Mullin, J. (1996). Dialogue structure and cooperative task performance in two CSCW environments. In J.H. Connolly & L. Pemberton (eds.). *Linguistics Concepts and Methods in CSCW* (Computer supported cooperative work). London: Springer-Verlag, 41-60.

O'Conaill, B. & Whittaker, S. (1993). Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8*, 389-428.

O'Conaill, B. & Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: a conversational approach. In Finn, K.E., Sellen, A.J. & Wilbur, S.B., Eds. *Video-mediated Communication*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Mally, C., Langton, S., Anderson, A., Doherty-Sneddon, G. & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers, 8*, 177-192.

Orestrom, B. (1983). *Turn-taking in English Conversation*. Lund: Sweden: Liber.

Plant, G. (1991). SYNTEX: Synthetic training exercises for hearing impaired adults. Sydney, Australia: National Acoustics Laboratories.

Preminger, J. & Levitt, H. (1998). Computer-Assisted Remote Transcription (CART): A tool to aid people who are deaf or hard of hearing in the workplace. *Volta Voices, 99(4)*, 219-230.

Robson, G.D. (1997). *Inside Captioning*. CyberDawg Pubiishing.

Rosen, K. & Yamplosky, S. (2000). Automatic speech recognition an a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication, 16*, 48-60.

Rosenfeld, H.M. (1987). Conversational control functions of nonverbal behavior. In A. Siegman & S. Feldstein (eds.) *Non-verbal Behavior and Communication* (pp. 563-601). Hillsdale, NJ: LEA.

Rutter (1987). *Communicating by telephone*. Oxford, UK: Pergamon

Sacks, H., Schegloff, E.A. & Jefferson, G.A. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696-735.

Schein, J.D. & Delk, M.T. (1974). *The Deaf Population of the United States*. Silver Spring, MD: National Association of the Deaf.

Steinfeld, A. (1999). The benefit of real-time captioning in a mainstream classroom as measured by working memory. *The Volta Review, 100(1)*, 29-45.

Siegel S., Castellan Jr., N.J. (1988). *Non-parametric Statistics for the Behavioral Sciences* (2nd Edition). ISBN 0-07-100326-6: 174-83.

Stuckless, R. (1994). Developments in real-time speech-to-text communication for people with impaired hearing. In Ross, M. (Ed.), *Communication Access for Persons with Hearing Loss*. Baltimore MD: York Press, 197-226.

Stuckless, R. (2000). Feasibility of verbatim shadowing and automatic speech recognition to transcribe speech into text in the college classroom for students who are deaf to read in real time. *National Technical Institute for the Deaf* (unpublished manuscript).

Stuckless, R. (Ed.) (1997). *Frank W. Lovejoy Symposium on Applications of Automatic Speech Recognitions with Deaf and Hard of Hearing people*. Rochester: Rochester, NY: Rochester Institute of Technology. http://www.rit.edu/~ewcncp/Lovejoy.html

Stuckless, R. (Ed.) (1998). Recognition means more than getting the words right: beyond accuracy to readability. *Speech Technology Magazine*, October/November.

Tannen. D. (1984). The pragmatics of cross-cultural communication. *Applied Linguistics, 5 (3)* 189-195.

Trychin, S. (1994). Helping people cope with hearing loss. In J. Greer Clark & N.F. Martin (Eds.), *Effective Counseling in Audiology* (pp. 247-277). Englewood Cliffs, NJ: Prentice-Hall.

Turn, R. (1974). The use of Speech for Man-machine Communication. RAND Report-1386-ARPA. Santa Monica, CA: RAND Corp.

Tye-Murray, N., Purdy, S.C., & Woodworth, G.G. (1992). Reported use of communication strategies by SHHH members: client, talker, and situational variables. *Journal of Speech and Hearing Research, 35*, 708-717.

Tye-Murray, N. & Witt, S. (1996). Conversational moves and conversational styles of adult cochlear-implant users. *Journal of the Academy of Rehabilitative Audiology, 29*, 11-25.

Tye-Murray, N., Witt, S. & Schum, L. (1995). Effect of talker familiarity on communication breakdown in conversations with adult cochlear implant users. *Ear and Hearing*, 16(5), 459-469.

Tye-Murray, N., Witt, S. & Schum, L. & Sobaski, C. (1994). Communication breakdowns: partner contingencies and partner reactions. *Journal of the Academy of Rehabilitative Audiology, 27*, 107-133.

Virvan, B. (1991). You don't have to hate meetings – Try computer-assisted notetaking. *SHHH Journal*, February.

Walton, A. & Doherty-Sneddon, G. (2000). Friends, familiars, and strangers: communicating with young intellectually-impaired adults. Presentations at Conference in Seattle, WA.

Weintraub, M., Taussig, K. Hunicke-Smith, K., & Snodgrass, A. (1996). The effect of speaking style on LVCSR Performance. *Proc. Intl. Conf. on Spoken Language Processing*, Addendum, 16-19, Philadelphia, PA.

Woodcock, K. & Aguayo, M. (1999). The duality of communication as a barrier between people who hear and people who don't. From M. Prosser and K. Sitaram (Eds.) *The Fourth Rochester Conference on Human Rights and Responsibilities: Communicating Among Nations and Peoples.*

Woodcock, K. (1992). Print interpreting for deafened adults. 1992 ALDA Reader. Chicago: Association of Late Deafened Adults.

Woodcock, K. (1997). Ergonomics and automatic speech recognition for deaf and hard-of-hearing people. *Technology and Disability, 7*, 147-164. http://www.isc.rit.edu/~klweie/asr.htm

Yarrow, M & Waxler, C (1979). Observing interaction: Confrontation with methodology. In (Cairns, R., Ed.), *The Analysis Of Social Interactions: Methods, Issues and Illustrations.* [pp. 37-66], Hillsdale: Lawrence Earlbaum Associates, Inc.

Youdelman, K.S. (2001). Computer-assisted notetaking in the classroom. *Volta Voices, 8(1),* 8-10.